

①⑨ RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

①⑪ N° de publication :

2 765 588

(à n'utiliser que pour les  
commandes de reproduction)

②① N° d'enregistrement national :

97 08816

⑤① Int Cl<sup>6</sup> : C 12 N 15/48, C 07 K 14/15, C 07 H 21/00, A 61 K 48/  
00, C 12 Q 1/68

⑫

DEMANDE DE BREVET D'INVENTION

A1

②② Date de dépôt : 07.07.97.

③⑦ Priorité :

④③ Date de mise à la disposition du public de la  
demande : 08.01.99 Bulletin 99/01.

⑤⑥ Liste des documents cités dans le rapport de  
recherche préliminaire : *Se reporter à la fin du  
présent fascicule*

⑥⑦ Références à d'autres documents nationaux  
apparentés :

⑦① Demandeur(s) : BIO MERIEUX SOCIETE ANONYME  
— FR.

⑦② Inventeur(s) :

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : GERMAIN ET MAUREAU.

⑤④ MATÉRIEL NUCLEIQUE RETROVIRAL ET FRAGMENTS NUCLEOTIDIQUES NOTAMMENT ASSOCIES A LA  
SCLEROSE EN PLAQUES ET/OU LA POLYARTHRITE RHUMATOÏDE, A DES FINS DE DIAGNOSTIC,  
PROPHYLACTIQUES ET THERAPEUTIQUES.

⑤⑦ Matériel nucléaire, à l'état isolé ou purifié, et fragment  
nucléotidique, comprenant une séquence nucléotidique  
choisie dans le groupe qui consiste en (i) les séquences  
SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ  
ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID  
NO: 141 et SEQ ID NO: 142; (ii) les séquences complémen-  
taires des séquences (i); et (iii) les séquences équivalentes  
aux séquences (i) ou (ii), en particulier les séquences pré-  
sentant pour toute suite de 100 monomères contigus, au  
moins 50 %, et préférentiellement au moins 70 % d'homolo-  
gie avec respectivement les séquences (i) ou (ii), et utilis-  
ations pour détecter un rétrovirus associé à la sclérose en  
plaques et/ ou à la polyarthrite rhumatoïde.

FR 2 765 588 - A1



La Sclérose en Plaques (SEP) est une maladie démyélinisante du système nerveux central (SNC) dont la cause complète reste encore inconnue.

De nombreux travaux ont étayé l'hypothèse d'une 5 étiologie virale de la maladie, mais aucun des virus connus testés ne s'est avéré être l'agent causal recherché: une revue des virus recherchés depuis des années dans la SEP a été faite par E. Norrby et R.T. Johnson.

10 Récemment, un rétrovirus, différent des rétrovirus humains connus a été isolé chez des patients atteints de SEP. Les auteurs ont aussi pu montrer que ce rétrovirus 15 pouvait être transmis in vitro, que des patients atteints de SEP produisaient des anticorps susceptibles de reconnaître des protéines associées à l'infection des cellules leptoméningées par ce rétrovirus, et que 20 l'expression de ce dernier pouvait être fortement stimulée par les gènes immédiats-précoces de certains herpesvirus.

Tous ces résultats plaident en faveur du rôle dans 25 la SEP d'au moins un rétrovirus inconnu ou d'un virus ayant une activité transcriptase inverse (RT) détectable selon la méthode publiée par H. Perron et qualifiée d'activité "RT de type LM7".

Les travaux de la Demanderesse ont permis 30 d'obtenir deux lignées continues de cellules infectées par des isolats naturels provenant de deux patients différents atteints de SEP, par un procédé de culture tel que décrit dans le document WO-A-93 20188, dont le contenu est incorporé par référence à la présente description. Ces 35 deux lignées dérivées de cellules de plexus-choroïdes humains, dénommées LM7PC et PLI-2 ont été déposées à l'E.C.A.C.C. respectivement le 22 juillet 1992 et le 8 janvier 1993, sous les numéros 92 072201 et 93 010817, conformément aux dispositions du Traité de Budapest. Par ailleurs, les isolats viraux possédant une activité RT de type LM7 ont également été déposés à l'E.C.A.C.C. sous la

dénomination globale de "souches". La "souche" ou isolat hébergé par la lignée PLI-2, dénommée POL-2, a été déposée auprès de l'E.C.A.C.C. le 22 juillet 1992 sous le n° V92072202. La "souche" ou isolat hébergé par la lignée LM7PC, dénommée MS7PG, a été déposée auprès de l'E.C.A.C.C. le 8 janvier 1993 sous le n° V93010816.

A partir des cultures et des isolats précités, caractérisés par des critères biologiques et morphologiques, on s'est ensuite attaché à caractériser le matériel nucléique associé aux particules virales produites dans ces cultures.

Les portions de génome déjà caractérisées ont été utilisées pour mettre au point des tests de détection moléculaire du génome viral et des tests immunosérologiques, utilisant les séquences d'acides aminés codées par les séquences nucléotidiques du génome viral, pour détecter la réponse immunitaire dirigée contre des épitopes associés à l'infection et/ou l'expression virale.

Ces outils ont déjà permis de confirmer une association entre la SEP et l'expression des séquences identifiées dans les brevets cités plus loin. Cependant, le système viral découvert par la Demanderesse, s'apparente à un système rétroviral complexe. En effet, les séquences retrouvées encapsidées dans les particules virales extracellulaires produites par les différentes cultures de cellules de patients atteints de SEP, montrent clairement qu'il y a co-encapsidation de génomes rétroviraux apparentés, mais différents du génome rétroviral "sauvage" qui produit les particules virales infectantes. Ce phénomène a été observé entre des rétrovirus réplicatifs et des rétrovirus endogènes appartenant à la même famille, voire même hétérologues. La notion de rétrovirus endogène est très importante dans le contexte de notre découverte car, dans le cas de MSRV-1, on a observé que des séquences rétrovirales endogènes comprenant des séquences homologues au génome MSRV-1,

existent dans l'ADN humain normal. L'existence d'éléments rétroviraux endogènes (ERV) apparentés à MSRV-1 par tout ou partie de leur génome, explique le fait que l'expression du rétrovirus MSRV-1 dans les cellules  
5 humaines puisse interagir avec des séquences endogènes proches. Ces interactions sont retrouvées dans le cas de rétrovirus endogènes pathogènes et/ou infectieux (par exemple certaines souches écotropes du Murine Leukaemia virus), dans le cas de rétrovirus exogènes dont la  
10 séquence nucléotidique peut être retrouvée partiellement ou en totalité, sous forme d'ERVs, dans le génome de l'animal hôte (ex. virus exogène de la tumeur mammaire de la souris transmis par le lait). Ces interactions consistent principalement en (i) une transactivation ou  
15 co-activation d'ERVs par le rétrovirus répliatif, (ii) une encapsidation "illégitime" d'ARN apparentés d'ERVs, ou d'ERVs -voire d'ARN cellulaires- possédant simplement des séquences d'encapsidation compatibles, dans les particules rétrovirales produites par l'expression de la souche  
20 répliatif, parfois transmissibles et parfois avec une pathogénicité propre, et (iii) des recombinaisons plus ou moins importantes entre les génomes co-encapsidés, notamment dans les phases de transcription inverse, qui conduisent à la formation de génomes hybrides, parfois  
25 transmissibles et parfois avec une pathogénicité propre.

Ainsi, (i) différentes séquences apparentées à MSRV-1 ont été retrouvées dans les particules virales purifiées; (ii) l'analyse moléculaire des différentes régions du génome rétroviral MSRV-1 doit être faite en  
30 analysant systématiquement les séquences co-encapsidées, interférantes et/ou recombinées qui sont générées par l'infection et/ou l'expression de MSRV-1, de plus, certains clones peuvent avoir des parties de séquences défectives produites par la répliation rétrovirale et les  
35 erreurs de matrice et/ou de transcription de la transcriptase inverse; (iii) les familles de séquences

apparentées à une même région génomique rétrovirale sont les supports d'une détection diagnostique globale qui peut être optimisée par l'identification de régions invariables parmi les clones exprimés et par l'identification de  
5 trames de lectures responsables de la production de polypeptides antigéniques et/ou pathogènes qui peuvent n'être produits que par une partie, voire un seul, des clones exprimés et dans ces conditions, l'analyse systématique des clones exprimés dans une région d'un gène  
10 donné permet d'évaluer la fréquence de variation et/ou de recombinaison du génome MSRV-1 dans cette région et de définir les séquences optimales pour les applications, notamment diagnostiques; (iv) la pathologie provoquée par un rétrovirus tel que MSRV-1 peut être un effet direct de  
15 son expression et des protéines ou peptides produits de ce fait, mais aussi un effet de l'activation, de l'encapsidation, de la recombinaison de génomes apparentés ou hétérologues et des protéines ou peptides produits de ces faits ; ainsi ces génomes associés à l'expression de  
20 et/ou l'infection par MSRV-1 sont-ils une partie intégrante de la pathogénicité potentielle de ce virus et donc, constituent des supports de détection diagnostique et des cibles thérapeutiques particulières. De même, tout agent associé à, ou, co-facteur de ces interactions  
25 responsables de la pathogénie en cause, tel que MSRV-2 ou le facteur gliotoxique décrit dans la demande de brevet publiée sous le N° FR-2 716 198, peut participer à l'élaboration d'une stratégie globale et très efficace de diagnostic, de pronostic, de suivi thérapeutique et/ou de  
30 thérapeutique intégrée de la SEP notamment, mais aussi de toute autre maladie associée aux mêmes agents.

Dans ce contexte, on a fait une découverte parallèle dans une autre maladie autoimmune, la polyarthrite rhumatoïde (PR), qui a été décrite dans la  
35 demande de brevet français déposée sous le N°95 02960. Cette découverte montre que, en appliquant des approches

méthodologiques similaires à celles qui furent utilisées dans les travaux de la Demanderesse sur la SEP, on a pu identifier un rétrovirus exprimé dans la PR qui partage les séquences décrites pour MSRV-1 dans la SEP et aussi, 5 la co-existence d'une séquence associée MSRV-2 également décrite dans la SEP. En ce qui concerne MSRV-1, les séquences détectées communément dans la SEP et la PR, concernent les gènes *pol* et *gag*. En l'état actuel des connaissances, on peut associer les séquences *gag* et *pol* 10 décrites aux souches MSRV-1 exprimées dans ces deux maladies.

La présente demande de brevet a pour objet différents résultats, supplémentaires par rapport à ceux déjà protégés par les demandes de brevet français :

- 15 - N° 92 04322 du 03.04.1992, publiée sous le N° 2 689 519;
- N° 92 13447 du 03.11.1992, publiée sous le N° 2 689 521;
- N° 92 13443 du 03.11.1992, publiée sous le N° 2 689 520;
- N° 94 01529 du 04.02.1994, publiée sous le N° 2 715 936;
- N° 94 01531 du 04.02.1994, publiée sous le N° 2 715 939;
- 20 - N° 94 01530 du 04.02.1994, publiée sous le N° 2 715 936;
- N° 94 01532 du 04.02.1994, publiée sous le N° 2 715 937;
- N° 94 14322 du 24.11.1994, publiée sous le N° 2 727 428;
- N° 94 15810 du 23.12.1994, publiée sous le N° 2 728 585;
- et
- 25 - la demande de brevet WO-97/06260.

La présente invention concerne tout d'abord un matériel nucléique, qui peut consister en un matériel rétroviral, à l'état isolé ou purifié, pouvant être appréhendé ou caractérisé de différentes manières :

- 30 - il comprend une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences  
SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,  
SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,  
SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences  
35 complémentaires aux séquences (i) ; et (iii) les séquences  
équivalentes aux séquences (i) ou (ii), en particulier les

séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii);

- 5           - il code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113,           SEQ ID N°115,           SEQ ID NO: 118, 10   SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137;
- son gène pol comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 112, SEQ ID NO: 124 et leurs séquences complémentaires;
- 15           - l'extrémité 5' de son gène pol commence au nucléotide 1419 de SEQ ID NO: 130;
- son gène pol code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % 20   d'homologie, avec la séquence peptidique SEQ ID NO: 113;
- l'extrémité 3' de son gène gag finit au nucléotide 1418 de SEQ ID NO: 130;
- son gène env comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le 25   groupe qui consiste en SEQ ID NO: 117, et ses séquences complémentaires;
- son gène env comprend une séquence nucléotidique qui commence au nucléotide 1 de SEQ ID NO: 117 et finit au nucléotide au nucléotide 233 de SEQ ID NO: 114;
- 30           - son gène env code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence SEQ ID NO: 118;
- la région U3R de son LTR 3' comprend une 35   séquence nucléotidique qui se termine au nucléotide 617 de SEQ ID NO: 114;

- la région RU5 de son LTR 5' comprend une séquence nucléotidique qui commence au nucléotide 755 de SEQ ID NO: 120 et finit au nucléotide 337 de SEQ ID NO: 141 ou SEQ ID NO: 142;

5 - un matériel nucléique rétroviral comprenant une séquence qui commence au nucléotide 755 de SEQ ID NO: 120 et qui se termine au nucléotide 617 de SEQ ID NO: 114;

- le matériel nucléique rétroviral tel que défini précédemment est en particulier associé à au moins une  
10 maladie auto-immune telle que la sclérose en plaques ou la polyarthrite rhumatoïde.

L'invention concerne aussi un fragment nucléotidique qui répond au moins à l'une des définitions suivantes :

15 - il comprend ou consiste en une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii)  
20 les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec  
25 respectivement les séquences (i) ou (ii);

- il comprend ou consiste en une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une  
30 séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

D'autres objets de la présente invention sont les suivants :

35 - une sonde nucléique pour la détection d'un rétrovirus associé à la sclérose en plaques et/ou la



polyarthrite rhumatoïde, susceptible de s'hybrider spécifiquement sur tout fragment précédemment défini et appartenant au génome dudit rétrovirus; elle possède avantageusement de 10 à 100 nucléotides, de préférence de  
5 10 à 30 nucléotides;

- une amorce pour l'amplification par polymérisation d'un ARN ou d'un ADN d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, qui comprend une séquence nucléotidique  
10 identique ou équivalente à au moins une partie de la séquence nucléotidique d'un fragment défini précédemment, notamment une séquence nucléotidique présentant pour toute suite de 10 monomères contigus, au moins 50 %, de préférence au moins 70 % d'homologie avec au moins ladite  
15 partie dudit fragment ; de préférence la séquence nucléotidique d'une amorce de l'invention est choisie parmi SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, et  
20 SEQ ID NO: 133;

- un ARN ou ADN, et notamment vecteur de réplication et/ou d'expression, comprenant un fragment génomique du matériel nucléique ou un fragment défini précédemment;

25 - un peptide codé par tout cadre de lecture ouvert appartenant à un fragment nucléotidique défini précédemment, notamment un polypeptide, par exemple oligopeptide formant ou comprenant un déterminant antigénique reconnu par des sera de patients infectés par  
30 le virus MSRV-1, et/ou chez lesquels le virus MSRV-1 a été réactivé; un peptide préférentiel comprend une séquence identique, partiellement ou totalement, ou équivalente à une séquence choisie parmi SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et  
35 SEQ ID NO: 137;

- une composition diagnostique, prophylactique, ou thérapeutique, notamment pour inhiber l'expression d'au moins un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, comprenant un fragment  
5 nucléotidique défini précédemment;

- un procédé pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, dans un échantillon biologique, comprenant les étapes consistant à mettre en contact un ARN et/ou un ADN présumé  
10 appartenir ou provenant dudit rétrovirus, ou leur ARN et/ou ADN complémentaire, avec une composition comprenant un fragment nucléotidique défini ci-dessus.

Avant de détailler l'invention, différents termes utilisés dans la description et les revendications sont à  
15 présent définis :

- par souche ou isolat, on entend toute fraction biologique infectante et/ou pathogène, contenant par exemple des virus et/ou des bactéries et/ou des parasites, générant un pouvoir pathogène et/ou antigénique, hébergée  
20 par une culture ou un hôte vivant ; à titre d'exemple, une souche virale selon la définition précédente peut contenir un agent co-infectant, par exemple un protiste pathogène,

- le terme "MSRV" utilisé dans la présente description désigne tout agent pathogène et/ou infectant, associé à la SEP, notamment une espèce virale, les souches atténuées de ladite espèce virale, ou les particules défectives interférentes ou contenant des génomes co-encapsidés ou encore des génomes recombinés avec une partie du génome MSRV-1, dérivées de cette espèce. Il est  
25 connu que les virus et particulièrement les virus contenant de l'ARN ont une variabilité, consécutive notamment à des taux relativement élevés de mutation spontanée, dont il sera tenu compte ci-après pour définir la notion d'équivalence,  
30

- par virus humain, on entend un virus susceptible d'infecter ou d'être hébergé par l'être humain,

- compte tenu de toutes les variations et/ou recombinaison naturelles ou induites, pouvant être rencontrées dans la pratique de la présente invention, les objets de cette dernière, définis ci-dessus et dans les  
5 revendications, ont été exprimés en comprenant les équivalents ou dérivés des différents matériels biologiques définis ci-après, notamment des séquences homologues nucléotidiques ou peptidiques,

- le variant d'un virus ou d'un agent pathogène  
10 et/ou infectant selon l'invention, comprend au moins un antigène reconnu par au moins un anticorps dirigé contre au moins un antigène correspondant dudit virus et/ou dudit agent pathogène et/ou infectant, et/ou un génome dont toute partie est détectée par au moins une sonde  
15 d'hybridation, et/ou au moins une amorce d'amplification nucléotidique spécifique dudit virus et/ou agent pathogène et/ou infectant, dans des conditions d'hybridation déterminées bien connues de l'homme de l'art,

- selon l'invention, un fragment nucléotidique ou  
20 un oligonucléotide ou un polynucléotide est un enchaînement de monomères, ou un biopolymère, caractérisé par la séquence informationnelle des acides nucléiques naturels, susceptible de s'hybrider à tout autre fragment nucléotidique dans des conditions prédéterminées,  
25 l'enchaînement pouvant contenir des monomères de structures chimiques différentes et être obtenu à partir d'une molécule d'acide nucléique naturelle et/ou par recombinaison génétique et/ou par synthèse chimique ; un fragment nucléotidique peut être identique à un fragment  
30 génomique du virus MSRV-1 considéré par la présente invention, notamment un gène de ce dernier, par exemple pol ou env dans le cas dudit virus ;

- ainsi un monomère peut être un nucléotide naturel d'acide nucléique, dont les éléments constitutifs  
35 sont un sucre, un groupement phosphate et une base azotée ; dans l'ARN le sucre est le ribose, dans l'ADN le sucre est

le désoxy-2-ribose; selon qu'il s'agit de l'ADN ou l'ARN, la base azotée est choisie parmi l'adénine, la guanine, l'uracile, la cytosine, la thymine; ou le nucléotide peut être modifié dans l'un au moins des trois éléments  
5 constitutifs ; à titre d'exemple, la modification peut intervenir au niveau des bases, générant des bases modifiées telles que l'inosine, la méthyl-5-désoxycytidine, la désoxyuridine, la diméthylamino-5-désoxyuridine, la diamino-2,6-purine, la  
10 bromo-5-désoxyuridine et toute autre base modifiée favorisant l'hybridation; au niveau du sucre, la modification peut consister dans le remplacement d'au moins un désoxyribose par un polyamide, et au niveau du groupement phosphate, la modification peut consister dans  
15 son remplacement par des esters, notamment choisis parmi les esters de diphosphate, d'alkyl et arylphosphonate et de phosphorothioate,

- par "séquence informationnelle", on entend toute suite ordonnée de monomères, dont la nature chimique et  
20 l'ordre dans un sens de référence, constituent ou non une information fonctionnelle de même qualité que celle des acides nucléiques naturels,

- par hybridation, on entend le processus au cours duquel, dans des conditions opératoires appropriées, deux  
25 fragments nucléotidiques, ayant des séquences suffisamment complémentaires, s'apparient pour former une structure complexe, notamment double ou triple, de préférence sous forme d'hélice,

- une sonde comprend un fragment nucléotidique  
30 synthétisé par voie chimique ou obtenu par digestion ou coupure enzymatique d'un fragment nucléotidique plus long, comprenant au moins six monomères, avantageusement de 10 à 100 monomères, de préférence 10 à 30 monomères, et possédant une spécificité d'hybridation dans des  
35 conditions déterminées ; de préférence, une sonde possédant moins de 10 monomères n'est pas utilisée seule,

mais l'est en présence d'autres sondes de taille aussi courte ou non ; dans certaines conditions particulières, il peut être utile d'utiliser des sondes de taille supérieure à 100 monomères ; une sonde peut notamment être  
5 utilisée à des fins de diagnostic et il s'agira par exemple de sondes de capture et/ou de détection,

- la sonde de capture peut être immobilisée sur un support solide par tout moyen approprié, c'est-à-dire directement ou indirectement, par exemple par covalence ou  
10 adsorption passive,

- la sonde de détection peut être marquée au moyen d'un marqueur choisi notamment parmi les isotopes radioactifs, des enzymes notamment choisis parmi la peroxydase et la phosphatase alcaline et ceux susceptibles  
15 d'hydrolyser un substrat chromogène, fluorigène ou luminescent, des composés chimiques chromophores, des composés chromogènes, fluorigènes ou luminescents, des analogues de bases nucléotidiques, et la biotine,

- les sondes utilisées à des fins de diagnostic de  
20 l'invention peuvent être mises en oeuvre dans toutes les techniques d'hybridation connues, et notamment les techniques dites "DOT-BLOT", "SOUTHERN BLOT", "NORTHERN BLOT" qui est une technique identique à la technique "SOUTHERN BLOT" mais qui utilise de l'ARN comme cible, la  
25 technique SANDWICH ; avantageusement, on utilise la technique SANDWICH dans la présente invention, comprenant une sonde de capture spécifique et/ou une sonde de détection spécifique, étant entendu que la sonde de capture et la sonde de détection doivent présenter une  
30 séquence nucléotidique au moins partiellement différente,

- toute sonde selon la présente invention peut s'hybrider in vivo ou in vitro sur l'ARN et/ou sur l'ADN, pour bloquer les phénomènes de réplication, notamment traduction et/ou transcription, et/ou pour dégrader ledit  
35 ADN et/ou ARN,

- une amorce est une sonde comprenant au moins six monomères, et avantageusement de 10 à 30 monomères, possédant une spécificité d'hybridation dans des conditions déterminées, pour l'initiation d'une polymérisation enzymatique, par exemple dans une technique d'amplification telle que la PCR (Polymerase Chain Reaction), dans un procédé d'élongation, tel que le séquençage, dans une méthode de transcription inverse ou analogue,

10 - deux séquences nucléotidiques ou peptidiques sont dites équivalentes ou dérivées l'une par rapport à l'autre, ou par rapport à une séquence de référence, si fonctionnellement les biopolymères correspondants peuvent jouer sensiblement le même rôle, sans être identiques, 15 vis-à-vis de l'application ou utilisation considérée, ou dans la technique dans laquelle elles interviennent ; sont notamment équivalentes deux séquences obtenues du fait de la variabilité naturelle, notamment mutation spontanée de l'espèce à partir de laquelle elles ont été identifiées, 20 ou induite, ainsi que deux séquences homologues, l'homologie étant définie ci-après,

- par "variabilité", on entend toute modification, spontanée ou induite d'une séquence, notamment par substitution, et/ou insertion, et/ou délétion de 25 nucléotides et/ou de fragments nucléotidiques, et/ou extension et/ou raccourcissement de la séquence à l'une au moins des extrémités; une variabilité non naturelle peut résulter des techniques de génie génétique utilisées, par exemple du choix des amorces de synthèse dégénérées ou 30 non, retenues pour amplifier un acide nucléique; cette variabilité peut se traduire par des modifications de toute séquence de départ, considérée comme référence, et pouvant être exprimées par un degré d'homologie par rapport à ladite séquence de référence,

35 - l'homologie caractérise le degré d'identité de deux fragments nucléotidiques ou peptidiques comparés ;

elle se mesure par le pourcentage d'identité qui est notamment déterminé par comparaison directe de séquences nucléotidiques ou peptidiques, par rapport à des séquences nucléotidiques ou peptidiques de référence,

5           - tout fragment nucléotidique est dit équivalent ou dérivé d'un fragment de référence, s'il présente une séquence nucléotidique équivalente à la séquence du fragment de référence ; selon la définition précédente, sont notamment équivalents à un fragment nucléotidique de  
10 référence :

(a) tout fragment susceptible de s'hybrider au moins partiellement avec le complément du fragment de référence,

(b) tout fragment dont l'alignement avec le  
15 fragment de référence conduit à mettre en évidence des bases contigues identiques, en nombre plus important qu'avec tout autre fragment provenant d'un autre groupe taxonomique,

(c) tout fragment résultant ou pouvant résulter de  
20 la variabilité naturelle de l'espèce, à partir de laquelle il est obtenu,

(d) tout fragment pouvant résulter des techniques de génie génétique appliquées au fragment de référence,

(e) tout fragment, comportant au moins huit  
25 nucléotides contigus, codant pour un peptide homologue ou identique au peptide codé par le fragment de référence,

(f) tout fragment différent du fragment de référence, par insertion, délétion, substitution d'au moins un monomère, extension, ou raccourcissement à l'une  
30 au moins de ses extrémités ; par exemple, tout fragment correspondant au fragment de référence, flanqué à l'une au moins de ses extrémités par une séquence nucléotidique ne codant pas pour un polypeptide,

- par polypeptide, on entend notamment tout  
35 peptide d'au moins deux acides aminés, notamment oligopeptide, protéine, extrait, séparé, ou

substantiellement isolé ou synthétisé, par l'intervention de la main de l'homme, notamment ceux obtenus par synthèse chimique, ou par expression dans un organisme recombinant,

- par polypeptide codé de manière partielle par un  
5 fragment nucléotidique, on entend un polypeptide présentant au moins trois acides aminés codés par au moins neuf monomères contigus compris dans ledit fragment nucléotidique,

- un acide aminé est dit analogue à un autre acide  
10 aminé, lorsque leurs caractéristiques physico-chimiques respectives, telles que polarité, hydrophobicité, et/ou basicité, et/ou acidité, et/ou neutralité, sont sensiblement les mêmes ; ainsi, une leucine est analogue à une isoleucine.

- tout polypeptide est dit équivalent ou dérivé  
15 d'un polypeptide de référence, si les polypeptides comparés ont sensiblement les mêmes propriétés, et notamment les mêmes propriétés antigéniques, immunologiques, enzymologiques et/ou de reconnaissance  
20 moléculaire ; est notamment équivalent à un polypeptide de référence :

(a) tout polypeptide possédant une séquence dont au moins un acide aminé a été substitué par un acide aminé analogue,

25 (b) tout polypeptide ayant une séquence peptidique équivalente, obtenue par variation naturelle ou induite dudit polypeptide de référence, et/ou du fragment nucléotidique codant pour ledit polypeptide,

(c) un mimotope dudit polypeptide de référence,

30 (d) tout polypeptide dans la séquence duquel un ou plusieurs acides aminés de la série L sont remplacés par un acide aminé de la série D, et vice versa,

(e) tout polypeptide dans la séquence duquel on a introduit une modification des chaînes latérales des  
35 acides aminés, telle que par exemple une acétylation des



fonctions amines, une carboxylation des fonctions thiol, une estérification des fonctions carboxyliques,

(f) tout polypeptide dans la séquence duquel une ou des liaisons peptidiques ont été modifiées, comme par exemple les liaisons carba, rétro, inverso, rétro-inverso, réduites, et méthylène-oxy,

(g) tout polypeptide dont au moins un antigène est reconnu par anticorps dirigé contre un polypeptide de référence,

10           - le pourcentage d'identité caractérisant l'homologie de deux fragments peptidiques comparés est selon la présente invention d'au moins 50% et de préférence au moins 70 %.

Etant donné qu'un virus possédant une activité  
15 enzymatique transcriptase inverse peut être génétiquement caractérisé aussi bien sous forme d'ARN que d'ADN, il sera fait mention aussi bien de l'ADN que de l'ARN viral pour caractériser les séquences relatives à un virus possédant une telle activité transcriptase inverse, dit MSRV-1 selon  
20 la présente description.

Les expressions d'ordre utilisées dans la présente description et les revendications, telles que "première séquence nucléotidique" ne sont pas retenues pour exprimer un ordre particulier, mais pour définir plus clairement  
25 l'invention.

Par détection d'une substance ou agent, on entend ci-après aussi bien une identification, qu'une quantification, ou une séparation ou isolement de ladite substance ou dudit agent.

30           L'invention sera mieux comprise à la lecture de la description détaillée qui va suivre faite en référence aux figures annexées dans lesquelles :

La Figure 1 représente la structure générale de l'ADN proviral et l'ARN génomique de MSRV-1.

35           La Figure 2 représente la séquence nucléotidique du clone dénommé CL6-5' (SEQ ID NO: 112) et trois trames

de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 3 représente la séquence nucléotidique du clone dénommé CL6-3' (SEQ ID NO: 114) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 4 représente la séquence nucléotidique du clone dénommé C15 (SEQ ID NO: 117) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 5 représente la séquence nucléotidique du clone dénommé 5M6 (SEQ ID NO: 120) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 6 représente la séquence nucléotidique du clone dénommé CL2 (SEQ ID NO: 130) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 7 représente trois trames de lecture potentielles en acides aminés exprimées par pET28C-clone 2 et figurant sous la séquence nucléotidique.

La Figure 8 représente trois trames de lecture potentielles en acides aminés exprimées par pET21C-clone 2 et figurant sous la séquence nucléotidique.

La Figure 9 représente la séquence nucléotidique du clone dénommé LB13 (SEQ ID NO: 141) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 10 représente la séquence nucléotidique du clone dénommé LA15 (SEQ ID NO: 142) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 11 représente la séquence nucléotidique du clone dénommé LB16 (SEQ ID NO: 124) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

EXEMPLE 1 : OBTENTION D'UNE REGION CL6-5' CODANT POUR L'EXTREMITÉ N-TERMINALE DE L'INTEGRASE ET D'UNE REGION CL6-3' CONTENANT LA SEQUENCE 3' TERMINALE DU GENOME MSRV-1

5

Une 3'RACE a été effectuée sur de l'ARN total extrait de plasma d'un patient atteint de SEP. Un plasma témoin sain, traité sous les mêmes conditions, a été utilisé comme contrôle négatif. La synthèse de cDNA a été réalisée avec une amorce oligo dT identifiée par SEQ ID NO: 68 (5' GAC TCG CTG CAG ATC GAT TTT TTT TTT TTT TTT T 3') et la transcriptase inverse "Expand<sup>TM</sup> RT" de Boehringer selon les conditions préconisées par la société. Une PCR a été effectuée avec l'enzyme Klentaq (Clontech) sous les conditions suivantes : 94°C 5 min puis 93°C 1 min, 58°C 1 min, 68°C 3 min pendant 40 cycles et 68°C pendant 8 min, avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

- 20 - amorce 5', identifiée par SEQ ID NO: 69  
5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3' ;
- amorce 3', identifiée par SEQ ID NO: 68

Une deuxième PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions expérimentales que celles utilisées lors de la première PCR, en utilisant 10 µl du produit d'amplification issu de la première PCR.

Amorces utilisées pour la PCR semi-nichée:

- 30 - amorce 5', identifiée par SEQ ID NO: 70  
5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3' ;
- amorce 3', identifiée par SEQ ID NO: 68

Les amorces SEQ ID NO: 69 et SEQ ID NO: 70 sont spécifiques de la région pol de MSRV-1.

35

Un produit d'amplification de 1,9Kb a été obtenu pour le plasma de patient SEP. Le fragment correspondant

n'a pas été observé pour le plasma témoin sain. Ce produit d'amplification a été cloné de la façon suivante :

l'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les 2 µl de solution d'ADN ont été

5 mélangés avec 5 µl d'eau distillée stérile, 1 µl d'un tampon de ligation 10 fois concentré "10X LIGATION BUFFER", 2 µl de "PCR<sup>TM</sup> VECTOR" (25 ng/ml) et 1 µl de "T4 DNA LIGASE". Ce mélange a été incubé une nuit à 12°C. Les étapes suivantes ont été réalisées conformément aux

10 instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La

15 préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le

20 séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur Sp6 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage

25 "PRISM<sup>TM</sup> Ready Reaction AmpliTaq® FS, DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

30 Le clone obtenu, contient une région CL6-5'codant pour l'extrémité N terminale de l'intégrase et une région CL6-3', correspondant à la région 3' terminale de MSRV-1 et permettant de définir la fin de l'enveloppe (234 pb) et les régions U3, R (401 pb) du rétrovirus MSRV1.

35 La région correspondant à l'extrémité N terminale de l'intégrase est représentée par sa séquence

nucléotidique (SEQ ID NO: 112) dans la figure 1. Les trois trames de lecture potentielles sont présentées par leur séquence aminoacide sous la séquence nucléotidique, et la séquence aminoacide de l'extrémité N-terminale de l'intégrase est identifiée par SEQ ID NO: 113.

La région Cl6-3' est représentée par sa séquence nucléotidique (SEQ ID NO: 114) dans la figure 3. Les trois trames de lecture potentielles sont présentées par leur séquence aminoacide sous la séquence nucléotidique. Une séquence aminoacide correspondant à l'extrémité C-terminale de la protéine env de MSRV-1 est identifiée par SEQ ID NO: 115.

EXEMPLE 2 : OBTENTION DU CLONE C15 CONTENANT LA REGION CODANT POUR UNE PARTIE DE L'ENVELOPPE DU RETROVIRUS MSRV-1

Une RT-PCR a été effectuée sur de l'ARN total extrait de virions concentrés par ultracentrifugation à partir d'un surnageant de culture de synoviocytes provenant d'un patient PR. La synthèse de cDNA a été réalisée avec une amorce oligo dT et la transcriptase inverse "Expand<sup>TM</sup> RT" de Boehringer selon les conditions préconisées par la société. Une PCR a été effectuée avec l'Expand<sup>TM</sup> Long Template PCR System (Boehringer) sous les conditions suivantes : 94°C 5 min puis 93°C 1 min, 60°C 1 min, 68°C 3 min pendant 40 cycles et 68°C pendant 8 min et avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

- amorce 5', identifiée par SEQ ID NO: 69
- 5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3' ;
- amorce 3', identifiée par SEQ ID NO: 116
- 5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

Une deuxième PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions expérimentales que celles

utilisées lors de la première PCR (sauf que 30 cycles ont été réalisés au lieu de 40), en utilisant 10 µl du produit d'amplification issu de la première PCR.

Amorces utilisées pour la PCR semi-nichée:

- 5 - amorce 5', identifiée par SEQ ID NO: 70  
5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3' ;
- amorce 3', identifiée par SEQ ID NO: 116

Les amorces SEQ ID NO: 69 et SEQ ID NO: 70 sont spécifiques de la région pol de MSRV-1. L'amorce SEQ ID  
10 NO: 116 est spécifique de la séquence FBd13 (aussi dénommé B13) et est localisée dans la région env conservée parmi les oncorétrovirus.

Un produit d'amplification de 1932 pb a été obtenu et cloné de la façon suivante :

- 15 l'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été
- 20 repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides
- 25 possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur SP6 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au
- 30 séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) et le
- 35 séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone C15 obtenu, contient une région correspondant à la région de l'enveloppe de MSRV-1, de 1481 pb.

La région env du clone C15 est représentée par sa  
5 séquence nucléotidique (SEQ ID NO: 117) dans la figure 5.  
Les trois trames de lecture potentielles de ce clone sont  
présentées par leur séquence aminoacide sous la séquence  
nucléotidique. La trame de lecture correspondant à une  
protéine env structurale MSRV-1 est identifiée par  
10 SEQ ID NO: 118.

EXEMPLE 3 : OBTENTION D'UN CLONE 5M6 CONTENANT LES  
SEQUENCES DE LA REGION 3' TERMINALE DE L'ENVELOPPE,  
SUIVIES DES SEQUENCES U3,R,U5 DE TYPE PROVIRAL MSRV-1.

15

Une PCR monodirectionnelle a été effectuée sur de  
l'ADN extrait de lymphocytes B immortalisés en culture  
d'un patient PR. La PCR a été effectuée avec l'Expand™  
Long Template PCR System (Boehringer) sous les conditions  
20 suivantes : 94°C 3 min puis 93°C 1 min, 60°C 1 min, 68°C 3  
min pendant 10 cycles , puis 93°C 1 min, 60°C 1 min avec  
15 sec d'extension à chaque cycle, 68°C 3 min pendant 35  
cycles et 68°C pendant 7 min et avec un volume réactionnel  
final de 50 µl.

25

L'amorce utilisée pour la PCR identifiée par  
SEQ ID NO: 119 est 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT  
AAC CG 3' ;

L'amorces SEQ ID NO: 119 est spécifique de la région env  
du clone C15.

30

Un produit d'amplification de 1673 pb a été  
obtenu et cloné de la façon suivante :  
l'ADN amplifié a été inséré dans un plasmide à l'aide du  
kit TA Cloning®. Les différentes étapes ont été réalisées  
conformément aux instructions du kit TA Cloning®  
35 (Invitrogen). A la fin de la procédure, les colonies  
blanches de bactéries recombinantes (white) ont été

repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone 5M6 obtenu, contient une région correspondant à la région 3' de l'enveloppe de MSRV-1, de 492 pb suivi des régions U3, R et U5 (837 pb) de MSRV1.

Le clone 5M6 est représenté par sa séquence nucléotidique (SEQ ID NO: 120) dans la figure 7. Les trois trames de lecture potentielles de ce clone sont présentées par leur séquence aminoacide sous la séquence nucléotidique. La trame de lecture correspondant à l'extrémité C-terminale de la protéine env MSRV-1 est identifiée par SEQ ID NO: 121.

**EXEMPLE 4 : OBTENTION DU CLONE LB16 CONTENANT LA REGION CODANT L'INTEGRASE DU RETROVIRUS MSRV-1**

Une RT-PCR a été effectuée sur de l'ARN total traité à la DNaseI et extrait à partir d'un plexus choroïde provenant d'un patient SEP. La synthèse de cDNA a été réalisée avec une amorce oligo dT et la transcriptase inverse "Expand™ RT" de Boehringer selon les conditions préconisées par la société. Un contrôle "no RT" a été



effectué parallèlement sur le même matériel. Une PCR a été effectuée avec la Taq polymérase (Perkin Elmer) sous les conditions suivantes : 95°C 5 min puis 95°C 1 min, 55°C 1 min, 72°C 2 min pendant 35 cycles et 72°C pendant 8 min et  
5 avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

- amorce 5', identifiée par SEQ ID NO: 122

5' GGC ATT GAT AGC ACC CAT CAG 3' ;

- amorce 3', identifiée par SEQ ID NO: 123

10 5' CAT GTC ACC AGG GTG GAA TAG 3'

L'amorce SEQ ID NO: 122 est spécifique de la région pol de MSRV-1 et plus précisément similaire à la région intégrase décrite précédemment. L'amorce SEQ ID NO 123 a été définie sur des séquences des clones  
15 obtenus lors d'essais préalables.

Un produit d'amplification d'environ 760 pb a été obtenu uniquement dans l'essai avec RT et a été cloné de la façon suivante :

l'ADN amplifié a été inséré dans un plasmide à l'aide du  
20 kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été repiquées pour être cultivées et permettre l'extraction  
25 des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage  
30 du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode  
35 préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™

Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

5 Le clone LB16 obtenu, contient les séquences correspondant à l'intégrase. La séquence nucléotidique de ce clone est identifiée par SEQ ID NO: 124 sur la figure 11, trois trames de lecture sont déterminées.

10 EXEMPLE 5: OBTENTION D'UN CLONE 2, CL2, CONTENANT EN 3' UNE PARTIE HOMOLOGUE AU GENE POL, CORRESPONDANT AU GENE PROTEASE, ET AU GENE GAG (GM3) CORRESPONDANT A LA NUCLEOCAPSID, ET UNE NOUVELLE REGION 5'CODANTE, CORRESPONDANT AU GENE GAG PLUS SPECIFIQUEMENT LA MATRICE  
15 ET LA CAPSID de MSRV-1.

Une amplification par PCR a été effectuée sur de l'ARN total extrait à partir de 100 µl de plasma d'un patient atteint de SEP. Un témoin eau, traité sous les  
20 mêmes conditions, a été utilisé comme contrôle négatif. La synthèse de cDNA a été réalisée avec 300 pmole d'une amorce aléatoire (GIBCO-BRL, France) et la transcriptase inverse "Expand RT" (BOEHRINGER MANNHEIM, France) selon les conditions préconisées par la société. Une  
25 amplification par PCR ("polymerase chain reaction") a été effectuée avec l'enzyme Taq polymérase (Perkin Elmer, France) en utilisant 10 µl de cDNA sous les conditions suivantes: 94°C 2 min, 55°C 1 min, 72°C 2 min puis 94°C 1 min, 55°C 1 min, 72°C 2 min pendant 30 cycles et 72°C  
30 pendant 7 min et avec un volume réactionnel final de 50 µl.

Amorces utilisées pour l'amplification par PCR:

- amorce 5', identifiée par SEQ ID NO: 126

5' CGG ACA TCC AAA GTG ATG GGA AAC G 3' ;

35 - amorce 3', identifiée par SEQ ID NO: 127

5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

Une deuxième amplification par PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions expérimentales que celles utilisées lors de la première PCR, en utilisant 10  $\mu$ l du produit d'amplification issu de la première PCR.

Amorces utilisées pour l'amplification par PCR semi-nichée:

- 10 - amorce 5', identifiée par SEQ ID NO: 128  
5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3' ;
- amorce 3', identifiée par SEQ ID NO: 129  
5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

Les amorces SEQ ID NO: et SEQ ID NO: sont  
15 spécifiques de la région pol, clone G+E+A, plus spécifiquement la région E: position nucleotidique n° 423 à n° 448. Les amorces utilisées dans la région 5' ont été définies sur des séquences de clones obtenus lors d'essais préalables.

20 Un produit d'amplification de 1511 pb a été obtenu à partir de l'ARN extrait du plasma de patient SEP. Le fragment correspondant n'a pas été observé pour le témoin eau. Ce produit d'amplification a été cloné de la façon suivante.

25 L'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning<sup>TM</sup>. Les 2  $\mu$ l de solution d'ADN ont été mélangés avec 5  $\mu$ l d'eau distillée stérile, 1  $\mu$ l d'un tampon de ligation 10 fois concentré "10X LIGATION BUFFER", 2  $\mu$ l de "pCR<sup>TM</sup> VECTOR" (25 ng/ml) et 1  $\mu$ l de "T4  
30 DNA LIGASE". Ce mélange a été incubé une nuit à 14°C. Les étapes suivantes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). Le mélange a été étalé après transformation de la ligation dans des bactéries *E. coli* INV $\alpha$ F'. A la fin de la procédure, les  
35 colonies blanches de bactéries recombinantes ont été repiquées pour être cultivées et permettre l'extraction

des plasmides incorporés selon la procédure dite de "minipréparation d'ADN" (17). La préparation de plasmide de chaque colonie recombinante a été coupée par l'enzyme de restriction Eco RI et analysée sur gel d'agarose. Les  
5 plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La  
10 réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM<sup>TM</sup> Ready Reaction Amplitaq® FS, DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les  
15 appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone obtenu, dénommé CL2, contient une région C-terminale similaire à la région 5' terminale des clones G+E+A de MSRV-1, qui permet de définir la région C-  
20 terminale du gène gag et une nouvelle région correspondante à la région N-terminale du gène gag MSRV-1.

CL2 permet de définir une région de 1511 pb présentant une phase ouverte de lecture dans la région N-terminale de 1077 pb codante pour 359 acides aminés et une  
25 phase non-ouverte de lecture, de 454 pb, correspondant à la région C-terminale du gène gag MSRV-1.

La séquence nucléotidique de CL2 est identifiée par SEQ ID NO: 130. Elle est représentée à la figure XX3,1, avec les trames de lecture potentielles en  
30 aminoacide.

Le fragment de 1077 pb de CL2 codant pour 359 acides aminés a été amplifié par PCR avec l'enzyme *Pwo* (5U/ $\mu$ l) (Boehringer Mannheim, France) en utilisant 1  $\mu$ l de la minipréparation de l'ADN du clone 2 sous les conditions  
35 suivantes : 95°C 1 min, 60°C 1 min, 72°C 2 min pendant 25

cycles et avec un volume réactionnel final de 50  $\mu$ l à l'aide des amorces:

- amorce 5' (*Bam* HI), identifié par SEQ ID NO: 132  
5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer), et
  - 5 - amorce 3' (*Hind* III), identifié par SEQ ID NO: 133  
5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer)
- correspondant, respectivement, à la séquence nucléotidique du clone 2 en position -9 à 21 et 1066 à 1095.

Le fragment obtenu après PCR, a été linéarisé par  
10 *Bam* HI et *Hind*III et sous-cloné dans les vecteurs d'expression pET28C et pET21C (NOVAGEN) linéarisé par *Bam* HI et *Hind* III. Le séquençage de l'ADN du fragment de 1077 pb du clone 2 dans les deux vecteurs d'expression a été réalisé selon la méthode préconisée pour l'utilisation du  
15 kit de séquençage "PRISM<sup>TM</sup> Ready Reaction Amplitaq<sup>®</sup> FS, DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

20 L'expression de la séquence nucléotidique du fragment de 1077 pb du clone 2 par les vecteurs d'expression pET28C et pET21C sont identifiées par respectivement SEQ ID NO: 135 et SEQ ID NO: 137.

#### 25 EXEMPLE 6:      EXPRESSION DU CLONE 2 CHEZ *ESCHERICHIA COLI*

Les constructions pET28c-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb) synthétisent, dans la souche bactérienne BL21 (DE3), une protéine en fusion N- et C-  
30 terminale pour le vecteur pET28C et C-Terminale pour le vecteur pET21C avec 6 Histidines, de masse moléculaire apparente d'environ 45 kDa, mise en évidence par électrophorèse sur gel de polyacrylamide SDS-PAGE (SDS = Dodecyl Sulfate de Sodium) (Laemmli, 1970 (1)). La  
35 réactivité de la protéine a été mise en évidence vis à vis

d'un anticorps monoclonal anti-Histidine (DIANOVA) par la technique de Western blot (Towbin et al., 1979 (2)).

Les protéines recombinantes pET28c-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb) ont été visualisées en SDS-PAGE dans la fraction insoluble après digestion enzymatique des extraits bactériens avec 50  $\mu$ l de lysozyme (10 mg/ml) et lyse par ultrasons.

Les propriétés antigéniques des antigènes recombinants pET28C-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb) ont été testées par Western Blot () après solubilisation du culot bactérien avec 2% SDS et 50 mM  $\beta$ -mercaptoéthanol. Après incubation avec les sérums de patients atteints de sclérose en plaques, les sérums des témoins neurologiques et les sérums de témoins de centre de transfusion sanguine (CTS), les immunocomplexes ont été détectés à l'aide d'un sérum de chèvre anti-IgG et anti IgM humaines, couplé à la phosphatase alcaline.

Les résultats sont présentés dans le tableau ci-après.

TABLEAU

Réactivité de sérums atteints de sclérose en plaques et témoins avec la protéine recombinante MSRV-1 gag clone 2 (1077 pb) = pET21C-clone 2 (1077 pb) et pET28C-clone 2 (1077 pb)<sup>a</sup>

| MALADIE       | NOMBRE<br>D'INDIVIDUS TESTÉS | NOMBRE<br>D'INDIVIDUS POSITIFS |
|---------------|------------------------------|--------------------------------|
| SEP           | 15                           | 6<br>2(+++), 2(++), 2(+)       |
| TEMOINS       |                              |                                |
| NEUROLOGIQUES | 2                            | 1(+++)                         |
| TEMOINS       |                              |                                |
| SAINS (CTS)   | 22                           | 1(+/-)                         |

(a) Les bandelettes contenant 1,5  $\mu$ g d'antigène recombinant pET-gag clone 2 (1077 pb) présentent une

réactivité contre de sérums dilués au 1/100. L'interprétation de Western Blot est basée sur la présence ou absence d'une bande pET-gag clone 2 (1077 pb) spécifique sur les bandelettes. Des contrôles positifs et  
5 négatifs sont inclus dans chaque expérience.

Ces résultats montrent que, dans les conditions techniques utilisées, environ 40% des sérums humains atteints de sclérose en plaques testés réagissent avec les  
10 protéines recombinantes pET28C-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb). Une réactivité a été observée sur un témoin neurologique et il est intéressant de noter que les ARN extraits à partir de ce sérum, après l'étape de transcriptase inverse, sont aussi amplifiés par PCR  
15 dans la région pol. Ceci suggère que des personnes n'ayant pas déclaré une SEP peuvent également héberger et exprimer ce virus. Par contre, un témoin (donneur CTS) apparemment sain, possède des anticorps anti-gag (clone 2, 1077 pb). Ce qui est compatible avec une immunité acquise contre  
20 MSRV-1 en dehors d'une maladie autoimmune associée déclarée.

EXEMPLE 7: OBTENTION D'UN CLONE LB13 CONTENANT EN 3' UNE PARTIE HOMOLOGUE AU CLONE 2 CORRESPONDANT AU GENE GAG  
25 ET EN 5' UNE PARTIE HOMOLOGUE AU CLONE 5M6 CORRESPONDANT À LA RÉGION LTR U5.

Une RT-PCR ("reverse-transcriptase-polymerase chain reaction") a été effectuée à partir de l'ARN total  
30 extrait de virions, provenant de surnageants de cellules lymphocytaires B des patients atteints de sclérose en plaques, concentrés par ultracentrifugations. La synthèse de cDNA a été réalisée avec une amorce spécifique SEQ N° XXX et la transcriptase inverse "Expand<sup>TM</sup> RT" de  
35 BOEHRINGER MANNHEIM selon les conditions préconisées par la société.

Amorce utilisée pour la synthèse du cDNA, identifiée par SEQ ID NO: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

Une amplification par PCR a été réalisée avec la  
5 Taq polymérase (Perkin Elmer, France) sous les conditions suivantes: 94°C 1 min, 55°C 1 min, 72°C 2 min pendant 35 cycles et 72°C pendant 7 min et avec un volume réactionnel final de 100 µl.

Amorces utilisées pour l'amplification par PCR:

10 - amorce 5', identifiée par SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- amorce 3', identifiée par SEQ ID NO: 138

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

Une deuxième amplification par PCR dite " semi-  
15 nichée " a été réalisée avec une amorce 3' située à l'intérieur de la région déjà amplifiée. Cette deuxième amplification a été effectuée sous les mêmes conditions expérimentales que celles utilisées lors de la première amplification, en utilisant 10 µl du produit  
20 d'amplification issu de la première PCR.

Amorces utilisées pour l'amplification par PCR " semi-nichée ":

- amorce 5', identifiée par SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

25 - amorce 3', identifiée par SEQ ID NO: 140

5' CTA TGT CCT TTT GGA CTG TTT GGG T 3'

Les amorces SEQ ID NO: 138 et SEQ ID NO: 140 sont  
spécifiques de la région gag, clone 2 position  
nucléotidique n° 373-397 et n° 433-456. Les amorces  
30 utilisées dans la région 5' ont été définies sur des séquences des clones obtenus lors d'essais préalables.

Un produit d'amplification de 764 pb a été obtenu et cloné de la façon suivante:

L'ADN amplifié a été inséré dans un plasmide à  
35 l'aide du kit TA Cloning<sup>TM</sup>. Les 2 µl de solution d'ADN ont été mélangés avec 5 µl d'eau distillée stérile, 1 µl d'un



tampon de ligation 10 fois concentré "10X LIGATION BUFFER", 2  $\mu$ l de "pCR<sup>TM</sup> VECTOR" (25 ng/ml) et 1  $\mu$ l de "T4 DNA LIGASE". Ce mélange a été incubé une nuit à 14°C. Les étapes suivantes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). Le mélange a été étalé après transformation de la ligation dans des bactéries *E. coli* INV $\alpha$ F'. A la fin de la procédure, les colonies blanches de bactéries recombinantes ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "minipréparation d'ADN" (17). La préparation de plasmide de chaque colonie recombinante a été coupée par l'enzyme de restriction *Eco* RI et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM<sup>TM</sup> Ready Reaction Amplitaq® FS, DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone LB13 obtenu contient une région N-terminale de gène gag MSRV-1 homologue au clone 2 et un LTR correspondant à une partie de la région U5. Entre la région U5 et gag un site de fixation pour les ARN de transfert, le PBS "primer binding site" a été identifié.

La séquence nucléotidique du fragment de 764 pb du clone LB13 dans le plasmide "pCR<sup>TM</sup> vector" est représentée dans l'identificateur SEQ ID NO: 141.

Le site de fixation pour les ARN de transfert, présentant une séquence du type PBS tryptophane, a été

identifié en position nucléotidique n°342-359 du clone LB13.

5 Un autre clone, dénommé LA15 a été obtenu sur l'ARN total extrait de virions concentrés par ultracentrifugation à partir d'un surnageant de culture de synoviocytes provenant d'un patient atteint de polyarthrite rhumatoïde. La stratégie d'amplification et clonage du clone LA15 est exactement la même qui a été utilisée pour le clone LB13.

10 La séquence nucléotidique du clone LA15 qui est représentée dans l'identificateur SEQ ID NO: 142, est très similaire au clone LB13. Ceci suggère que le rétrovirus MSVR-1 détecté dans la sclérose en plaque présente des séquences similaires à celles rencontrées dans la  
15 polyarthrite rhumatoïde.

## BIBLIOGRAPHIE

- (1) Laemmli U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4.  
5 Nature. (1970). 227: 680-685.
- (2) Towbin H., Staehelin T. & Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some  
10 applications. Proc. Natl. Acad. Sci. USA. (1979). 76: 4350-4354.

## LISTE DE SEQUENCES

## (2) INFORMATIONS POUR LA SEQ ID NO: 68:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 5 (A) LONGUEUR: 34 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## 10 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 68:

GACTCGCTGC AGATCGATT TTTTTTTTTT TTTT

34

## (2) INFORMATIONS POUR LA SEQ ID NO: 69:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 15 (A) LONGUEUR: 30 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## 20 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 69:

GCCATCAAGC CACCCAAGAA CTCTTAAGT

30

## (2) INFORMATIONS POUR LA SEQ ID NO: 70:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 25 (A) LONGUEUR: 30 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## 30 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 70:

CCAATAGCCA GACCATTATA TACACTAATT

30

## (2) INFORMATIONS POUR LA SEQ ID NO: 112

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 310 paires de bases

(B) TYPE: nucléotidique

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 112:

GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC CCCAGTACTC 60  
 AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT CCAGATGGCT 120  
 10 AGCCACTGAG GAAGGAAAAA TACTTTCACC TGCAGCTAAC CAACAGAAAT TACTTAAAAC 180  
 CCTTCACCAA ACCTTCCACT TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT 240  
 TACTGGACCA GGCCTTTTCA AAACATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA 300  
 AAGAAATAAT 310

## 15 (2) INFORMATIONS POUR LA SEQ ID NO: 113

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 103 acides aminés

(B) TYPE: peptidique

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 113:

LIEGPLVWGNPLWETKPYQSAGKIEKETSQGHFTFLPSRWLATEEGKILSPAANQQKLLKTLHQTFHLGID  
 STHQMAKLLFTGPGFLFKTIKKIVRGCEVCQRNN

## 25 (2) INFORMATIONS POUR LA SEQ ID NO: 114

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 635 paires de bases

(B) TYPE: nucléotidique

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 114:

CCCTGTATCT TTAACCTCCT TGTTAAGTTT GTCTCTTCCA GAATCAAAAC TGTA AAACTA 60  
 CAAATTGTTT TTCAAATGGA GCACCAGATG GAGTCCATGA CTAAGATCCA CCGTGGACCC 120  
 CTGGACCGGC CTGCTAGCCC ATGCTCCGAT GTTAATGACA TTGAAGGCAC CCCTCCCGAG 180  
 35 GAAATCTCAA CTGCACAACC CTTACTATGC CCCAATTCAG CGGGAAGCAG TTAGAGCGGT 240  
 CATCAGCCAA CCTCCCCAAC AGCACTGGG TTTTCCTGTT GAGAGGGGGG ACTGAGAGAC 300

37

AGGACTAGCT GGATTCCTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG GAAGGTGACT 360  
 GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA CCCGACCAAT CAGAGAGCTC 420  
 ACTAAAATGC TAATTAGGCA AAAATAGGAG GTAAAGAAAT AGCCAATCAT CTATTGCCTG 480  
 AGAGCACAGC GGGAGGGACA AGGATCGGGA TATAAACCCA GGCATTGAG CCGGCAACGG 540  
 5 CAACCCCTT TGGGTCCCCT CCCTTTGTAT GSGCGCTCTG TTTTCACTCT ATTTCACTCT 600  
 ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635

## (2) INFORMATIONS POUR LA SEQ ID NO: 115

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 10 (A) LONGUEUR: 77 acides aminés  
 (B) TYPE: peptidique  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 115:

15 PCIFNLLVKFVSSRIKTVKLQIVLQMEHQMESMTKIHGPLDRPASPCSDVNDIEGTPPEEISTAQPLLC  
 PNSAGSS

## (2) INFORMATIONS POUR LA SEQ ID NO: 116

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 20 (A) LONGUEUR: 32 paires de bases  
 (B) TYPE: nucléotidique  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 116:

25 TGGGGTTCCA TTGTAAGAC CATCTGTAGC TT

32

## (2) INFORMATIONS POUR LA SEQ ID NO: 117

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 30 (A) LONGUEUR: 1481 paires de base  
 (B) TYPE: nucléotidique  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 117:

ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCTTT CGCTCTCACT 60  
 35 GCACCCCTC CATGCTGCTG TACAACCACT AGCTCCCTT ACCAAGAGTT TCTATGAAGA 120  
 ACGCGGCTTC CTGGAATAT TGATGCCCCA TCATATAGGA GTTTATCTAA GGGAACTCC 180

ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG 240  
 CATGCAAATA CTCATTATTG GACAGGGAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 300  
 GGAGCCACTG TCTGTTGGAC TTACTIONTACC CATACCAGTA TGTCTGATGG GGGTGGAAAT 360  
 CAAGGTCAGG CAAGAGAAAA ACAAGTAAAG GAAGCAATCT CCCAACTGAC CCGGGGACAT 420  
 5 AGCACCCCTA GCCCCTACAA AGGACTAGTT CTCTCAAAAC TACATGAAAC CCTCCGTACC 480  
 CATACTCGCC TGGTGAGCCT ATTTAATACC ACCCTCACTC GGCTCCATGA GGTCTCAGCC 540  
 CAAAACCCCTA CTAAGTGTG GATGTGCCTC CCCCTGCACT TCAGGCCATA CATTTCATC 600  
 CCTGTTCTTG AACAATGGAA CAAGTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 660  
 GGACCTCTTG TTTCCAATCT GGAAATAACC CATACCTCAA ACCTCACCTG TGTAATAATT 720  
 10 AGCAATACTA TAGACACAAC CAGCTCCCAA TGCATCAGGT GGGTAACACC TCCCACACGA 780  
 ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTTG 840  
 AATGGCTCTT CAGAATCTAT GTGCTTCTC TCATTCTTAG TGCCCCCTAT GACCATCTAC 900  
 ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG AGTACCCATT 960  
 CTCCTTTTGT TTATCAGAGC AGGAGTGCTA GGCAGACTAG GTACTGGCAT TGGCAGTATC 1020  
 15 ACAACCTCTA CTCAGTTCTA CTACAACTA TCTCAAGAAA TAAATGGTGA CATGGAACAG 1080  
 GTCAGTACT CCCTGCTCAC CTGCAAGAT CAAGTTAACT CCCTAGCAGC AGTAGTCCTT 1140  
 CAAAATCGAA GAGCTTTAGA CTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTGA 1200  
 GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA AGTTAAAGAA 1260  
 ATTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAA ACACCGAACG CTGGGGCCTC 1320  
 20 CTCAGCCAAT GGATGCCCTG GGTCTCTCCC TTCTTAGGAC CTCTAGCAGC TCTAATATTG 1380  
 TTACTCTCTT TTGGACCCTG TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT 1440  
 GAAGCTGTAA AGCTACAGAT GGTCTTACAA ATGGAACCCC A 1481

(2) INFORMATIONS POUR LA SEQ ID NO: 118

25 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 493 acides aminés

(B) TYPE: peptidique

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

30 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 118:

MALPYHTFLFTVLLPPFALTAPPPCCCTSSSPYQEFLEXRTLPGNIDAPSYRSLSKGNSTFTAHTHMPR  
 NCYNSATLCMHANTHYWTGKMINPSCPGGLGATVCWTFYFHTSMSDGGGIQGGQAREKQVKEAISQLTRGH  
 STSPYKGLVLSKLHETLRTHRLVSLFNNTLRLHEVSAQNPTNCWMCLPLHFRPYISIPVPEQWNNFS  
 TEINTTSVLVGPLVSNLEITHTSNLTVCVKSNTIDTSSQCIRWVTPPTRIVCLPSGIFVCGTSAYHCL  
 35 NGSSSESMCFSLFVPPMTIYTEQDLNHHVVPKPHNKRVPILPFVIRAGVLGRLGTGIGSITTSTQFYKYL  
 SQEINGDMEQVTDLSLVTLDQLNSLAADVVLQNRRLDLLTAKRGGTCLFLGEERCYYVNQSRIVTEKVKE

IRDRIQCRAEELQNTERWGLLSQWMPWVLPFLGPLAALILLLLFGPCIFNLLVKFVSSRIEAVKLQMLVQ  
MEP

(2) INFORMATIONS POUR LA SEQ ID NO: 119

5 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 32 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

10 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 119:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CG

32

(2) INFORMATIONS POUR LA SEQ ID NO: 120

(i) CARACTERISTIQUES DE LA SEQUENCE:

15 (A) LONGUEUR: 1329 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 120:

|    |   |      |
|----|---|------|
| 20 | TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAGA GGGGGAACCT GTTTATTTTT  | 60   |
|    | AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC ATTACTGAGA AAGTTAAAGA | 120  |
|    | AATTTGAGAT CGAATATAAT GTAGAGCAGA GGACCTTCAA AACACTGCAC CCTGGGGCCT | 180  |
|    | CCTCAGCCAA TGGATGCCCT GGACTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT | 240  |
|    | TTTACTCCTC TTTGGACCCT GTATCTTCAA CTTCTTGTT AAGTTTGTCT CTTCCAGAAT  | 300  |
| 25 | TGAAGCTGTA AAGCTACAAA TAGTCTTCA AATGGAACCC CAGATGCAGT CCATGACTAA  | 360  |
|    | AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC TCTGATGTTA ATGACATTGA | 420  |
|    | AGTCACCCCT CCCGAGGAAA TCTCAACTGC ACAACCCCTA CTACACTCCA ATTCAGTAGG | 480  |
|    | AAGCAGTTAG AGCAGTTGTC AGCCAACCTC CCCAACAGTA CTTGGGTTTT CCTGTTGAGA | 540  |
|    | GGGTGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT | 600  |
| 30 | ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG CTCACACCCG | 660  |
|    | ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAA CAGGAGGTAA AGCAATAGCC  | 720  |
|    | AATCATCTAT TGCCTGAGAG CACAGCGGGA AGGACAAGGA TTGGGATATA AACTCAGGCA | 780  |
|    | TTCAAGCCAG CAACAGCAAC CCCCTTTGGG TCCCCTCCCA TTGTATGGGA GCTCTGTTTT | 840  |
|    | CACTCTATTT CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG | 900  |
| 35 | CTCAAGCTGA GCTTTTGTTT GCCATCCACC ACTGCTGTTT GCCACCGTCA CAGACCCGCT | 960  |
|    | GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG TGCTCCTGAT CCAGCGAGGT | 1020 |



ACCCATG GCC ACTCCCGATC AGGCTAAAGG CTTGCCATTG TTCCTGCATG GCTAAGTGCC 1080  
 TGGGTTTGTCT CTAATAGAAC TGAACACTGG TCACTGGGTT CCATGGTTCT CTTCCATGAC 1140  
 CCACGGCTTC TAATAGAGCT ATAACACTCA CCGCATGGCC CAAGATTCCA TTCCTTGGTA 1200  
 TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTGAGGCT TGCCACCATT TGGGAAGTGG 1260  
 5 CCCACTGCCA TTTTGGTAGC GGCCACCAC CATCTTGGGA GCTGTGGGAG CAAGGATCCC 1320  
 CCAGTAACA 1329

(2) INFORMATIONS POUR LA SEQ ID NO: 121

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 10 (A) LONGUEUR: 162 acides aminés  
 (B) TYPE: peptide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 121:

15 QNRRALDLLTAKRGGTCLFLGEECCXYVNSQSGIITEKVKEIXDRIXCRAEDLQNTAPWGLLSQWMPWTLF  
 FLGPLAAIIFLLFGPCIFNFLVKFVSSRIEAVKLQIVLQMEPQMOSMTKIYRGPLDRPARLCSDVNDIE  
 VTPPEEISTAQPLLHSNSVGSS

(2) INFORMATIONS POUR LA SEQ ID NO: 122

20 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: paires de base  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

25 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 122:

GGCATTGATA GCACCCATCA G

21

(2) INFORMATIONS POUR LA SEQ ID NO: 123

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 30 (A) LONGUEUR: 21 paires de base  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 123:

35 CATGTCACCA GGGTGGAATA G

21

.41

## (2) INFORMATIONS POUR LA SEQ ID NO: 124

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 124:

10

15

## (2) INFORMATIONS POUR LA SEQ ID NO: 126

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 21 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 126:

CATGTCACCA GGGTGGAATA G

21

25

## (2) INFORMATIONS POUR LA SEQ ID NO: 127

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 26 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 127:

GGACAGGAAA GTAAGACTGA GAAGCC

26

## 35 (2) INFORMATIONS POUR LA SEQ ID NO: 128

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

5 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 128:

cf Exemple 5

(2) INFORMATIONS POUR LA SEQ ID NO: 129

(i) CARACTERISTIQUES DE LA SEQUENCE:

10 (A) LONGUEUR: paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 129:

15 cf Exemple 5

(2) INFORMATIONS POUR LA SEQ ID NO: 130

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 1511 paires de base

20 (B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 130:

|    |  |     |
|----|--|-----|
|    | CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG  | 60  |
| 25 | ATTTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG  | 120 |
|    | GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA  | 180 |
|    | GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACCTTC TTTTCATTAA GAGACAACCTC | 240 |
|    | ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA  | 300 |
|    | CCCCAGCGTC CCCTCCCOGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC  | 360 |
| 30 | GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCCCG  | 420 |
|    | ATTATGCCCC CTCCAAGCAG TGAGAGGAGG AGAATTGGGC CCAGCCAGAG TGCCTGTACC  | 480 |
|    | TTTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAATTCT CAGATAACCC   | 540 |
|    | TGACGGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA  | 600 |
|    | TATAATGTTA CTAATAATC AGACACTAAC CCCAAATGAG AGAAGTGCCG CTGTAACTGC   | 660 |
| 35 | AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAATAGGA TGACAACAGA  | 720 |
|    | GGAAAGAACA ACTCCACAG GCCAGCAGGC AGTTCCCAGT GTAGACCCTC ATTGGGACAC   | 780 |

AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTTGCTA ACTTGCGTGC TAGAAGGACT 840  
 GAGGAAAAC AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA 900  
 GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC 960  
 CCTGTCACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC 1020  
 5 AGCTGCAGAC ATTAGAAAAA ACTTCAAAG TCTGCCTTAG GCCCGGAGCA GAACTTAGAA 1080  
 ACCCTATTTA ACTTGGCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCCAAA 1140  
 CGGACAAAAC GGGATAAAAA AAAAAGGGGG GTTCCACTAC TTTAGTCATG GCCCTCAGGC 1200  
 AAGCAGACTT TGGAGGCTCT GCAAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260  
 CTGGCTTCCA GTGCGGTCTA CAAGGACACT TAAAAAAGA TTATCCAAGT AGAAATAAGC 1320  
 10 CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380  
 GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440  
 GCGCGGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCGCGGGTA TGTTTGACCA 1500  
 TTGAGAGCCA A 1511

15 (2) INFORMATIONS POUR LA SEQ ID NO: 131

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 347 acides aminés

(B) TYPE: peptide

(C) NOMBRE DE BRINS: simple

20 (D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 131:

LERILENWDQCDTQTLRKKRFIFFCSTAWPOYPLOGRETWLPEGSINYNIILQLDLFCRKEGKWSEVPYV  
 QTFPSLRDNSQLCKKCGLCPTGSPQSPPPYPSPVPSPTPSSTNKDPPLTQTVQKEIDKGVNNEPKSANIPR  
 LCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLKGFSDNPDGYIDVLQGLGQSFDLTWRDIMLLNQTLT  
 25 PNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTGQQA VPSVDPHWDTESEHGDWCHKHLLTCVLEGL  
 RKTRKKPMNYSMSTITQGKEENLTAFLDRLREALRKHTSLSPDSIEGQLILKDKFITQSAADIRKNPKS  
 LP

(2) INFORMATIONS POUR LA SEQ ID NO: 132

30 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 30 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

35 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 132:

TGCTGGAATT CGGGATCCTA GAACGTATTC

## (2) INFORMATIONS POUR LA SEQ ID NO: 133

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 30 paires de base
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 133:

AGTTCTGCTC CGAAGCTTAG GCAGACTTTT

30

## (2) INFORMATIONS POUR LA SEQ ID NO: 135

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: acides aminés
- (B) TYPE: peptide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 135:

MGSSHHHHHSSGLVPRGSHMASMTGGQQMGRILERILENWDQCDTQTLRKKRFIFFCSTAWPQYPLQGR  
ETWLPEGSINYNIILQLDLFCRKEGKWSEVPYVQTFSLRDNSQLCKKCGLCPTGSPQSPPPYPSVPSPT  
PSSTNKDPPLTQTVQKEIDKGVNNEPKSANIPRLCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLGKFS  
DNPBGYIDVLQGLGQSFDLTWRDIMLLLNQTLTPNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTG  
QQAVPSVDPHWDTESEHGDWCHKHLLTCVLEGLRKRKPKMNYSMSTITQKKEENLTAFLDRLREALRK  
HTSLSPDSIEGQLILKDKFITQSAADIRKNFKSLPKLAAALEHHHHHH

## 10 (2) INFORMATIONS POUR LA SEQ ID NO: 137

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: acides aminés

(B) TYPE: peptide

(C) NOMBRE DE BRINS: simple

15 (D) CONFIGURATION: linéaire

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 137:

MASMTGGQQMGRILERILENWDQCDTQTLRKKRFFFCSTAWPQYPLQGRETWLPEGSINYNIILQLDLF  
CRKEGKWSEVPYVQTFFSLRDNSQLCKKCLCPTGSPQSPPPYPSVPSPTPSSTNKDPPLTQTVQKEIDK  
GVNNEPKSANIPRLCPLQAVRGGEFGPARVPVPFSLSDLKQIKIDLGKFSDNPDGYIDVLQGLGQSFDLT  
20 WRDIMLLLNQTLTPNERSAAVTAAREFGDLWYLSQANNRMTTEERTTPTGQAVPSVDPHWDTESEHGDW  
CHKHLLTCVLEGLRKTRKKPMNYSMMSTITQGKEENLTAFLDRLREALRKHTSLSPDSIEGQLILKDKFI  
TQSAADIRKNFKSLPKLAAALEHHHHHHH

## (2) INFORMATIONS POUR LA SEQ ID NO: 138

## 25 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 25 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

## 30 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 138:

CTTGGAGGGT GCATAACCAG GGAAT

25

## (2) INFORMATIONS POUR LA SEQ ID NO: 139

## (i) CARACTERISTIQUES DE LA SEQUENCE:

35 (A) LONGUEUR: 20 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 139:

TGTCCGCTGT GCTCCTGATC

20

5

(2) INFORMATIONS POUR LA SEQ ID NO: 140

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 25 paires de base

(B) TYPE: nucléotide

10

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 140:

CTATGTCCTT TTGGACTGTT TGGGT

25

15 (2) INFORMATIONS POUR LA SEQ ID NO: 141

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 764 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

20

(D) CONFIGURATION: linéaire

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 141:

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATGCCC TCTCCCAATT GGGCTAAAGG 60

CTTGCCATTG TTCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120

TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180

25 CTGCATGGTC CAAGATTCCA TTCCTTGGA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240

ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGAAGC AGCCCGCCAC 300

TATCTTGGGA GCTCTGGGAG CAAGGACCCC AGGTAACAAT TTGGTGACCA CGAAGGGACC 360

TGAATCGCA ACCATGAAGG GATCTCCAAA GCAATTGGAA ATGTTCTCTC CAAGGCAAAA 420

ATGCCCCCTAA GATGTATTCT GGAGAATTGG GACCAATTTG ACCCTCAGAC AGTAAGAAAA 480

30 AAATGACTTA TATTCTTCTG CAGTACCGCC CTGGCCACGA TATCCTCTTC AAGGGGGAGA 540

AACCTGGCCT CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACCTGTTTTG 600

TAGAAAAGGA GGCAAATGGA GTGAAGTGCC ATATTTACAA ACTTTCTTTT CATTAAAAGA 660

CAACTCGCAA TTATGTTAAC ACTGTGATTT GTGTTCTTAC ACCGAAGCCC TCAGATTCTA 720

CTCCCCACCC CCGGCATCTC CCCTGAATCC CTCCCCAACT TATT 764

35

(2) INFORMATIONS POUR LA SEQ ID NO: 142

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 800 paires de base

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

5

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 142:

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60  
 CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120  
 TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180  
 10 CTGCATGGTC CAAGATTCCA TTCCTTGGA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240  
 ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGGAAGC GGCCCGCCAC 300  
 TATCTTGGA GCTCTGGGAG CAAGGACCCC CAGGTAACAA TTTGGTGACC ACGAAGGGAC 360  
 CTGAATCCGC AACCATGAAG GGATCTCCAA AGCAATTGGA AATGTTCTC CCAAGGCAAA 420  
 AATGCCCCTA AGATGTATTC TGGAGAATTG GGACCAATCT GACCCTCAGA CAGTAAGAAA 480  
 15 AAAAATGACT TATATTCTTC TGCAGTACCG CCTGGCCACG GATATCCTCT TCAAGGGGGA 540  
 GAAACCTGGC CTCCTGAGGG AAGTATAAAT TATAACACCA TCTTACAGCT AGACCTGTTT 600  
 TGTAGAAAAG GAGGCAAATG GAGTGAAGTG CCATATTTAC AAACCTTCTT TTCATTAAAA 660  
 GACAACTCGC AATTATGTAA ACAGTGTGAT TTGTGTCCTA CAGGAAGCCC TCAGATCTAC 720  
 CTCCTTACCC CGGCATCTCC CTGACTCCTT CCCCACTAA TAAGGACCCA CTTCAGCCCA 780  
 20 AACAGTCCAA AAGGACATAG 800



## REVENDICATIONS

1. Matériel nucléique, à l'état isolé ou purifié, comprenant une séquence nucléotidique choisie dans le  
5 groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux  
10 séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii).
2. Matériel nucléique, à l'état isolé ou purifié,  
15 codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118,  
20 SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.
3. Matériel nucléique rétroviral, dont le gène pol comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 112, SEQ ID NO: 124 et leurs  
25 séquences complémentaires.
4. Matériel nucléique rétroviral, dont l'extrémité 5' du gène pol commence au nucléotide 1419 de  
SEQ ID NO: 130.
5. Matériel nucléique rétroviral, dont le gène pol  
30 code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence peptidique SEQ ID NO: 113.
6. Matériel nucléique rétroviral, dont l'extrémité  
35 3' du gène gag finit au nucléotide 1418 de SEQ ID NO: 130.

7. Matériel nucléique rétroviral, dont le gène env comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 117, et ses séquences  
5 complémentaires.

8. Matériel nucléique rétroviral, dont le gène env comprend une séquence nucléotidique qui commence au nucléotide 1 de SEQ ID NO: 117 et finit au nucléotide au nucléotide 233 de SEQ ID NO: 114.

10 9. Matériel nucléique rétroviral, dont le gène env code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence SEQ ID NO: 118.

15 10. Matériel nucléique rétroviral dont la région U3R du LTR 3' comprend une séquence nucléotidique qui se termine au nucléotide 617 de SEQ ID NO: 114.

20 11. Matériel nucléique rétroviral dont la région RU5 du LTR 5' comprend une séquence nucléotidique qui commence au nucléotide 755 de SEQ ID NO: 120 et finit au nucléotide 337 de SEQ ID NO: 141 ou SEQ ID NO: 142.

12. Matériel nucléique rétroviral comprenant une séquence qui commence au nucléotide 755 de SEQ ID NO: 120 et qui se termine au nucléotide 617 de SEQ ID NO: 114.

25 13. Matériel nucléique rétroviral selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il est associé à au moins une maladie auto-immune telle que la sclérose en plaques ou la polyarthrite rhumatoïde.

30 14. Fragment nucléotidique comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii)  
35 les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en

particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii).

5 15. Fragment nucléotidique selon la revendication 14, consistant en une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences  
SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,  
SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,  
10 SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères  
15 70 % d'homologie avec respectivement les séquences (i) ou (ii).

16. Fragment nucléotidique comprenant une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins  
20 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en  
SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118,  
SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

17. Fragment nucléotidique selon la revendication  
25 16, consistant en une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113,  
30 SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121,  
SEQ ID NO: 135 et SEQ ID NO: 137.

18. Sonde nucléique pour la détection d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle est  
35 susceptible de s'hybrider spécifiquement sur tout fragment

selon l'une quelconque des revendications 14 à 17, appartenant au génome dudit rétrovirus.

19. Sonde selon la revendication 18, caractérisée en ce qu'elle possède de 10 à 100 nucléotides, de  
5 préférence de 10 à 30 nucléotides.

20. Amorce pour l'amplification par polymérisation d'un ARN ou d'un ADN d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle comprend une séquence nucléotidique  
10 identique ou équivalente à au moins une partie de la séquence nucléotidique d'un fragment selon l'une quelconque des revendications 8 à 11, notamment une séquence nucléotidique présentant pour toute suite de 10 monomères contigus, au moins 50 %, de préférence au moins  
15 70 % d'homologie avec au moins ladite partie dudit fragment.

21. Amorce selon la revendication 20, caractérisée en ce que sa séquence nucléotidique est choisie parmi  
SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122,  
20 SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127,  
SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, et  
SEQ ID NO: 133.

22. ARN ou ADN, et notamment vecteur de répllication et/ou d'expression, comprenant un fragment  
25 génomique du matériel nucléique selon l'une quelconque des revendications 1 à 7 ou un fragment selon l'une quelconque des revendications 14 à 17.

23. Peptide codé par tout cadre de lecture ouvert appartenant à un fragment nucléotidique selon l'une  
30 quelconque des revendications 14 à 17, notamment polypeptide, par exemple oligopeptide formant ou comprenant un déterminant antigénique reconnu par des sera de patients infectés par le virus MSRV-1, et/ou chez lesquels le virus MSRV-1 a été réactivé.

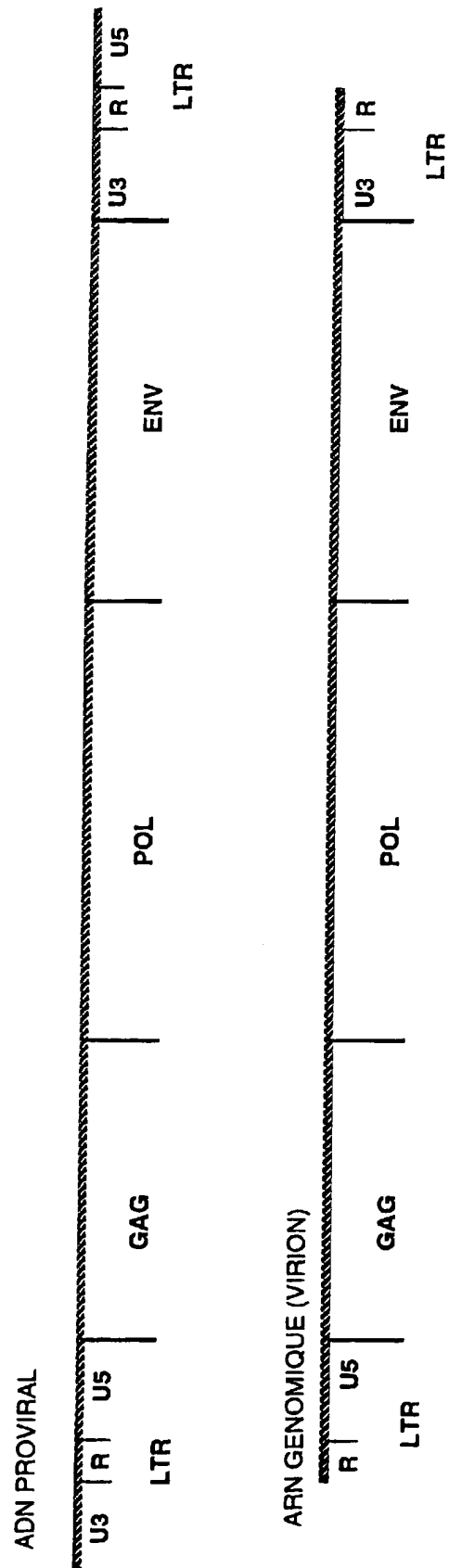
35 24. Peptide selon la revendication 23 comprenant une séquence identique, partiellement ou totalement, ou

équivalente à une séquence choisie parmi SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

25. Composition diagnostique, prophylactique, ou  
5 thérapeutique, notamment pour inhiber l'expression d'au moins un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, comprenant un fragment nucléotidique selon l'une quelconque des revendications 14 à 17.
- 10 26. Procédé pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, dans un échantillon biologique, caractérisé en ce que l'on met en contact un ARN et/ou un ADN présumé appartenir ou provenant dudit rétrovirus, ou leur ARN et/ou ADN  
15 complémentaire, avec une composition comprenant un fragment nucléotidique selon l'une quelconque des revendications 14 à 17.

1/27

FIG 1



2/27

FIG 2

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| GCTTATAGAA | GGACCCCTAG | TATGGGGTAA | TCCCTCTGG  | GAAACCAAGC | 50  |
| A Y R R    | T P S      | M G .      | S P L G    | N Q A      |     |
| L I E      | G P L V    | W G N      | P L W      | E T K P    |     |
| L . K      | D P .      | Y G V I    | P S G      | K P S      |     |
| CCAGTACTC  | AGCAGGAAAA | ATAGAATAGG | AAACCTCACA | AGGACATACT | 100 |
| P V L      | S R K N    | R I G      | N L T      | R T Y F    |     |
| Q Y S      | A G K      | I E .      | E T S Q    | G H T      |     |
| P S T Q    | Q E K      | . N R      | K P H K    | D I L      |     |
| TTCCTCCCT  | CCAGATGGCT | AGCCACTGAG | GAAGGAAAAA | TACTTTCAAC | 150 |
| P P L      | Q M A      | S H .      | G R K N    | T F T      |     |
| F L P S    | R W L      | A T E      | E G K I    | L S P      |     |
| S S P      | P D G .    | P L R      | K E K      | Y F H L    |     |
| TGCAGCTAAC | CAACAGAAAT | TACTTAAAAC | CCTTCACCAA | ACCTTCCACT | 200 |
| C S .      | P T E I    | T . N      | P S P N    | L P L      |     |
| A A N      | Q Q K L    | L K T      | L H Q      | T F H L    |     |
| Q L T      | N R N      | Y L K P    | F T K      | P S T      |     |
| TAGGCATTGA | TAGCACCCAT | CAGATGGCCA | AATTATTATT | TACTGGACCA | 250 |
| R H .      | . H P S    | D G Q      | I I I      | Y W T R    |     |
| G I D      | S T H      | Q M A K    | L L F      | T G P      |     |
| . A L I    | A P I      | R W P      | N Y Y L    | L D Q      |     |
| GGCCTTTTCA | AACTATCAA  | GAAGATAGTC | AGGGGCTGTG | AAGTGTGCCA | 300 |
| P F Q      | N Y Q      | E D S Q    | G L .      | S V P      |     |
| G L F K    | T I K      | K I V      | R G C E    | V C Q      |     |
| A F S      | K L S R    | R . S      | G A V      | K C A K    |     |
| AAGAAATAAT |            |            |            |            | 310 |
| K K .      |            |            |            |            |     |
| R N N      |            |            |            |            |     |
| E I        |            |            |            |            |     |

3/27

FIG 2 (suite)

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| CCCTGTATCT | TTAACTCTCT | TGTTAAGTTT | GTCTCTTCCA | GAATCAAAC  | 50  |
| P C I F    | N L L      | V K F      | V S S R    | I K T      |     |
| P V S      | L T S L    | L S L      | S L P      | E S K L    |     |
| L Y L      | . P P      | C . V C    | L F Q      | N Q N      |     |
|            |            |            |            |            |     |
| TGTAAACTA  | CAAATGTTC  | TTCAAATGGA | GCAACAGATG | GAGTCCATGA | 100 |
| V K L      | Q I V L    | Q M E      | H Q M      | E S M T    |     |
| . N Y      | K L F      | F K W S    | T R W      | S P .      |     |
| C K T T    | N C S      | S N G      | A P D G    | V H D      |     |
|            |            |            |            |            |     |
| CTAAGATCCA | COGTGGACCC | CTGGACCGGC | CTGCTAGCCC | ATGCTCCGAT | 150 |
| K I H      | R G P      | L D R P    | A S P      | C S D      |     |
| L R S T    | V D P      | W T G      | L L A H    | A P M      |     |
| . D P      | P W T P    | G P A C    | . P        | M L R C    |     |
|            |            |            |            |            |     |
| GTTAATGACA | TTGAAGGCAC | COCTCCCGAG | GAAATCTCAA | CTGCACAACC | 200 |
| V N D I    | E G T      | P P E      | E I S T    | A Q P      |     |
| L M T      | L K A P    | L P R      | K S Q      | L H N P    |     |
| . . H      | . R H      | P S R G    | N L N      | C T T      |     |
|            |            |            |            |            |     |
| OCTACTATGC | CCCAATTCAG | CGGGAAGCAG | TTAGAGCGGT | CATCAGCCAA | 250 |
| L L C      | P N S A    | G S S      | . S G      | H Q P T    |     |
| Y Y A      | P I Q      | R E A V    | R A V      | I S Q      |     |
| P T M P    | Q F S      | G K Q      | L E R S    | S A N      |     |
|            |            |            |            |            |     |
| CCGCCCAAC  | AGCACTTGGG | TTTTCTGT   | GAGAGGGGGG | ACTGAGAGAC | 300 |
| S P T      | A L G      | F S C      | . E G G    | L R D      |     |
| P P Q Q    | H L G      | F P V      | E R G D    | . E T      |     |
| L P N      | S T W V    | F L L      | R G G      | T E R Q    |     |
|            |            |            |            |            |     |
| AGGACTAGCT | GGATTTCCTA | GGCCAACGAA | GAATCCCTAA | GCCTAGCTGG | 350 |
| R T S W    | I S .      | A N E      | E S L S    | L A G      |     |
| G L A      | G F P R    | P T K      | N P .      | A . L G    |     |
| D . L      | D F L      | G Q R R    | I P K      | P S W      |     |



4/27

FIG 3

| 10            | 20            | 30          | 40         | 50         |     |
|---------------|---------------|-------------|------------|------------|-----|
| 1234567890    | 1234567890    | 1234567890  | 1234567890 | 1234567890 |     |
| GAAGGTGACT    | GCATCCACCT    | CTAAACATGG  | GGCTTGCAAC | TTAGCTCACA | 400 |
| K V T A S T S | K H G A C N   | L A H T     |            |            |     |
| R . L H P P   | L N M G       | L A T .     | L T        |            |     |
| E G D C I H L | . T W G L     | Q L S S H   |            |            |     |
| CCCGACCAAT    | CAGAGAGCTC    | ACTAAATGC   | TAATTAGGCA | AAAATAGGAG | 450 |
| R P I R E L   | T K M L       | I R Q K .   | E          |            |     |
| P D Q S E S S | L K C .       | L G K N R R |            |            |     |
| P T N Q R A H | . N A N .     | A K I G G   |            |            |     |
| GTAAAGAAAT    | AGCCAATCAT    | CTATTGCTTG  | AGAGCACAGC | GGGAGGGACA | 500 |
| V K K . P I I | Y C L R A Q R | E G Q       |            |            |     |
| . R N S Q S S | I A . E H S   | G R D K     |            |            |     |
| K E I A N H   | L L P E S T A | G G T       |            |            |     |
| AGGATCGGGA    | TATAAACCCA    | GGCATTGAG   | CCGGCAACGG | CAACCCCCCT | 550 |
| G S G Y K P R | H S S R Q R   | Q P P L     |            |            |     |
| D R D I N P   | G I R A G N G | N P L       |            |            |     |
| R I G I . T Q | A F E P A T A | T P F       |            |            |     |
| TGGGTCCCCCT   | CCCTTGTAT     | GGCGCTCTG   | TTTCACTCT  | ATTCACTCT  | 600 |
| G P L P L Y   | G R S V F T L | F H S       |            |            |     |
| W V P S L C M | G A L F S L Y | F T L       |            |            |     |
| G S P P F V W | A L C F H S   | I S L Y     |            |            |     |
| ATTAAATCTT    | GCAACTGAAA    | AAAAAAAAAA  | AAAAA      |            | 635 |
| I K S C N .   | K K K K K     |             |            |            |     |
| L N L A T E K | K K K K K     |             |            |            |     |
| . I L Q L K   | K K K K K     |             |            |            |     |

5/27

FIG 4

| 10   | 20         | 30         | 40         | 50         |     |
|--|------------|------------|------------|------------|-----|
| 1234567890   | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCTTT  |            |            |            |            | 50  |
| M A L P  | Y H T      | F L F      | T V L L    | P P F      |     |
| W P S  | L I I L    | F S L      | L F S      | Y P L S    |     |
| G P P  | L S Y      | F S L Y    | C S L      | T P F      |     |
| CGCTCTCACT GCACCCCTC CATGCTGCTG TACAACCACT AGCTCCCTT   |            |            |            |            | 100 |
| A L T  | A P P P    | C C C      | T T S      | S S P Y    |     |
| L S L  | H P L      | H A A V    | Q P V      | A P L      |     |
| R S H C  | T P S      | M L L      | Y N Q      | L P L      |     |
| ACCAAGAGTT TCTATGAAGA ACGGGCTTC CTGGAAATAT TGATGCCCCA  |            |            |            |            | 150 |
| Q E F L  | R T R L P  | G N I      | D A P      |            |     |
| T K S F  | Y E E      | R G F      | L E I L    | M P H      |     |
| P R V  | S M K N    | A A S      | W K Y      | C P I      |     |
| TCATATAGGA GTTATCTAA GGGAACTCC ACCTTCACTG CCCACCCCA    |            |            |            |            | 200 |
| S Y R S  | L S K      | G N S      | T F T A    | H T H      |     |
| H I G  | V Y L R    | E T P      | P S L      | P T P I    |     |
| I . E  | F I .      | G K L H    | L H C      | P H P      |     |
| TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG CATGCAAAIA |            |            |            |            | 250 |
| M P R  | N C Y N    | S A T      | L C M      | H A N T    |     |
| C P A  | T A I      | T L P L    | F A C      | M Q I      |     |
| Y A P Q  | L L .      | L C H      | S L H A    | C K Y      |     |
| CTCATTATTG GACAGGGAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT |            |            |            |            | 300 |
| H Y W  | T G K      | M I N P    | S C P      | G G L      |     |
| L I I G  | Q G K      | L I        | L V V L    | E D L      |     |
| S L L  | D R E N    | D . S      | L S        | W R T W    |     |
| GGAGCCACTG TCTGTTGGAC TTACTTCACC CATAACAGTA TGTCTGATGG |            |            |            |            | 350 |
| G A T V  | C W T      | Y F T      | H T S M    | S D G      |     |
| E P L  | S V G L    | T S P      | I P V      | C L M G    |     |
| S H C  | L L D      | L L H P    | Y Q Y      | V . W      |     |

6/27

FIG 4 (suite)

| 10            | 20            | 30          | 40         | 50          |      |
|---------------|---------------|-------------|------------|-------------|------|
| 1234567890    | 1234567890    | 1234567890  | 1234567890 | 1234567890  |      |
| ACCTCACTG     | TGTAAAATT     | AGCAATACTA  | TAGACACAAC | CAGCTCCCAA  | 750  |
| L T C V K F   | S N T I       | D T T S S Q |            |             |      |
| T S P V .     | N L A I L .   | T Q P A P N |            |             |      |
| P H L C K I . | Q Y Y R H N   | Q L P M     |            |             |      |
| TGCATCAGGT    | GGGTAAACACC   | TCCCACACGA  | ATAGTCTGCC | TACCCCTCAGG | 800  |
| C I R W V T P | P T R I V C L | P S G       |            |             |      |
| A S G G .     | H L P H E .   | S A Y P Q E |            |             |      |
| H Q V G N T   | S H T N S L P | T L R       |            |             |      |
| AATATTTTTT    | GTCGTGGTA     | CCTCAGCCTA  | TCATTGTTTG | AATGGCTCTT  | 850  |
| I F F V C G T | S A Y H C L   | N G S S     |            |             |      |
| Y F L S V V   | P Q P I I V . | M A L       |            |             |      |
| N I F C L W Y | L S L S L F E | W L F       |            |             |      |
| CAGAATCTAT    | GIGCTTCCTC    | TCATTCTTAG  | TGCCCCCTAT | GACCATCTAC  | 900  |
| E S M C F L   | S F L V P P M | T I Y       |            |             |      |
| Q N L C A S S | H S . C P L . | P S T       |            |             |      |
| R I Y V L P L | I L S A P Y   | D H L H     |            |             |      |
| ACTGAACAAG    | ATTTATACAA    | TCATGTGGTA  | OCTAAGCCCC | ACAACAAAAG  | 950  |
| T E Q D L Y N | H V V P K P H | N K R       |            |             |      |
| L N K I Y T I | M S Y L S P   | T T K E     |            |             |      |
| . T R F I Q   | S C R T .     | A P Q Q K   |            |             |      |
| AGTACCCATT    | CTTCCTTTTG    | TTATCAGAGC  | AGGAGTGCTA | GGCAGACTAG  | 1000 |
| V P I L P F V | I R A G V L   | G R L G     |            |             |      |
| Y P F F L L   | L S E Q E C . | A D .       |            |             |      |
| S T H S S F C | Y Q S R S A R | Q T R       |            |             |      |
| GTACTGGCAT    | TGGCAGTATC    | ACAACCTCTA  | CTCAGTTCTA | CTACAAACTA  | 1050 |
| T G I G S I   | T T S T Q F Y | Y K L       |            |             |      |
| V L A L A V S | Q P L L S S T | T N Y       |            |             |      |
| Y W H W Q Y H | N L Y S V L   | L Q T I     |            |             |      |

7/27

FIG 4 (suite)

| 10         | 20           | 30         | 40          | 50          |      |
|------------|--------------|------------|-------------|-------------|------|
| 1234567890 | 1234567890   | 1234567890 | 1234567890  | 1234567890  |      |
| TCTCAAGAAA | TAAATGGTGA   | CATGGAACAG | GTCACGTGACT | CCCTGGTCAC  | 1100 |
| S Q E I    | N G D        | M E Q      | V T D S     | L V T       |      |
| L K K      | . M V T      | W N R      | S L T       | P W S P     |      |
| S R N      | K W .        | H G T G    | H . L       | P G H       |      |
| CTTGCAAGAT | CAACTTAACT   | CCCTAGCAGC | AGTAGTCCTT  | CAAAATCGAA  | 1150 |
| L Q D      | Q L N S      | L A A      | V V L       | Q N R R     |      |
| C K I      | N L T        | P . Q Q    | . S F       | K I E       |      |
| L A R S    | T . L        | P S S      | S S P S     | K S K       |      |
| GAGCTTTAGA | CTTGCTAACC   | GCCAAAAGAG | GGGGAACCTG  | TTTATTTTIA  | 1200 |
| A L D      | L L T        | A K R G    | G T C       | L F L       |      |
| E L .      | T C .        | P P K E    | G E P V     | Y F .       |      |
| S F R      | L A N R      | Q K R      | G N L       | F I F R     |      |
| GGAGAAGAAC | GCTGTATTATTA | TGTTAATCAA | TCCAGAATTG  | TCACTGAGAA  | 1250 |
| G E E R    | C Y Y        | V N Q      | S R I V     | T E K       |      |
| E K N      | A V I M      | L I N      | P E L       | S L R K     |      |
| R R T      | L L L        | C . S I    | Q N C       | H . E       |      |
| AGTTAAAGAA | ATTGAGATC    | GAATACAATG | TAGAGCAGAG  | GACCTTCAAA  | 1300 |
| V K E      | I R D R      | I Q C      | R A E       | E L Q N     |      |
| L K K      | F E I        | E Y N V    | E Q R       | S F K       |      |
| S . R N    | S R S        | N T M      | . S R G     | A S K       |      |
| ACACCGAACG | CTGGGGCCTC   | CTCAGCCAAT | GGATGCCCTG  | GGTTCCTCCCC | 1350 |
| T E R      | W G L        | L S Q W    | M P W       | V L P       |      |
| T P N A    | G A S        | S A N      | G C P G     | F S P       |      |
| H R T      | L G P P      | Q P M      | D A L       | G S P L     |      |
| TTCTTAGGAC | CTCTAGCAGC   | TCTAATATTG | TTACTCCTCT  | TTGGACCCCTG | 1400 |
| F L G P    | L A A        | L I L      | L L L F     | G P C       |      |
| S . D      | L . Q L      | . Y C      | Y S S       | L D P V     |      |
| L R T      | S S S        | S N I V    | T P L       | W T L       |      |

8/27

## FIG 4 (suite)

---

| 10   | 20         | 30         | 40         | 50         |      |   |   |   |   |   |   |   |   |   |   |      |  |
|--|------------|------------|------------|------------|------|---|---|---|---|---|---|---|---|---|---|------|--|
| 1234567890   | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |   |   |   |   |   |   |   |   |   |   |      |  |
| TATCTTTAAC CTCCTTGTTA AGTTTGTC TCACAGAATT GAAGCTGTAA |            |            |            |            | 1450 |   |   |   |   |   |   |   |   |   |   |      |  |
| I  | F          | N          | L          | L          | V    | K | F | V | S | S | R | I | E | A | V | K    |  |
| S  | L          | T          | S          | L          | L    | S | L | S | L | P | E | L | K | L | . |      |  |
| Y  | L          | .          | P          | P          | C    | . | V | C | L | F | Q | N | . | S | C | K    |  |
| AGCTACAGAT GGTCTTACAA ATGGAACCCC A                   |            |            |            |            |      |   |   |   |   |   |   |   |   |   |   | 1481 |  |
| L  | Q          | M          | V          | L          | Q    | M | E | P |   |   |   |   |   |   |   |      |  |
| S  | Y          | R          | W          | S          | Y    | K | W | N | P |   |   |   |   |   |   |      |  |
| A  | T          | D          | G          | L          | T    | N | G | T | P |   |   |   |   |   |   |      |  |

9/27

FIG 5

| 10   | 20         | 30         | 40         | 50         |     |
|--|------------|------------|------------|------------|-----|
| 1234567890   | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT |            |            |            |            | 50  |
| S K S K  | S F R      | L A N      | R Q K R    | G N L      |     |
| Q N R  | R A L D    | L L T      | A K R      | G G T C    |     |
| K I E  | E L .      | T C .      | P P K E    | G E P      |     |
| GTTTATTTTT AGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC  |            |            |            |            | 100 |
| F I F  | R G R M    | L L V      | C . S      | I W N H    |     |
| L F L  | G E E      | C C .      | Y V N Q    | S G I      |     |
| V Y F .  | G K N      | A V S      | M L I N    | L E S      |     |
| ATTACTGAGA AAGTTAAAGA AATTIGAGAT CGAATATAAT GTAGAGCAGA |            |            |            |            | 150 |
| Y . E  | S . R      | N L R S    | N I M .    | S R        |     |
| I T E K  | V K E      | I . D      | R I . C    | R A E      |     |
| L L R  | K L K K    | F E I      | E Y N      | V E Q R    |     |
| GGACCTTCAA AACACTGCAC CCTGGGGCCT CCTCAGCCAA TGGATGCCCT |            |            |            |            | 200 |
| G P S K  | H C T      | L G P      | P Q P M    | D A L      |     |
| D L Q  | N T A P    | W G L      | L S Q      | W M P W    |     |
| T F K  | T L H      | P G A S    | S A N      | G C P      |     |
| GGACTCTOCC CTTCCTAGGA CCTCTAGCAG CTATAATATT TTTACTCCTC |            |            |            |            | 250 |
| D S P  | L L R T    | S S S      | Y N I      | F T P L    |     |
| T L P  | F L G      | P L A A    | I I F      | L L L      |     |
| G L S P  | S . D      | L . Q      | L . Y F    | Y S S      |     |
| TTTGGACCCCT GTATCTTCAA CTTCCTTGTT AAGTTTGICT CTTCAGAAT |            |            |            |            | 300 |
| W T L  | Y L Q      | L P C .    | V C L      | F Q N      |     |
| F G P C  | I F N      | F L V      | K F V S    | S R I      |     |
| L D P  | V S S T    | S L L      | S L S      | L P E L    |     |
| TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT |            |            |            |            | 350 |
| . S C K  | A T N      | S S S      | N G T P    | D A V      |     |
| E A V  | K L Q I    | V L Q      | M E P      | Q M Q S    |     |
| K L .  | S Y K .    | F F K      | W N P      | R C S      |     |

10/27

FIG 5 (suite)

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| CCATGACTAA | AATCTACCGT | GGACCCCTGG | ACCGGCCTGC | TAGACTATGC | 400 |
| H D .      | N L P W    | T P G      | P A C      | . T M L    |     |
| M T K      | I Y R      | G P L D    | R P A      | R L C      |     |
| P . L K    | S T V      | D P W      | T G L L    | D Y A      |     |
| TCTGATGTTA | ATGACATTGA | AGTCACCCCT | CCCGAGGAAA | TCTCAACTGC | 450 |
| . C .      | . H .      | S H P S    | R G N      | L N C      |     |
| S D V N    | D I E      | V T P      | P E E I    | S T A      |     |
| L M L      | M T L K    | S P L      | P R K      | S Q L H    |     |
| ACAACCCCTA | CTACACTCCA | ATTGAGTAGG | AAGCAGTTAG | AGCAGTTGTC | 500 |
| T T P T    | T L Q      | F S R      | K Q L E    | Q L S      |     |
| Q P L      | L H S N    | S V G      | S S .      | S S C Q    |     |
| N P Y      | Y T P      | I Q .      | E A V R    | A V V      |     |
| AGCCAACCTC | CCCAACAGTA | CTTGGGTTTT | OCTGTTGAGA | GGGIGGACTG | 550 |
| A N L      | P N S T    | W V F      | L L R      | G W T E    |     |
| P T S      | P T V      | L G F S    | C .        | E G G L    |     |
| S Q P P    | Q Q Y      | L G F      | P V E R    | V D .      |     |
| AGAGACAGGA | CTAGCTGGAT | TTCCTAGGCT | GACTAAGAAT | CCCAAGCCT  | 600 |
| R Q D      | . L D      | F L G .    | L R I      | P K P      |     |
| R D R T    | S W I      | S .        | A D .      | E S X S L  |     |
| E T G      | L A G F    | P R L      | T K N      | P X A X    |     |
| ANCTGGGAAG | GTGACCGCAT | CCATCTTTAA | ACATGGGGCT | TGCAACTTAG | 650 |
| X W E G    | D R I      | H L .      | T W G L    | Q L S      |     |
| X G K      | V T A S    | I F K      | H G A      | C N L A    |     |
| L G R      | . P H      | P S L N    | M G L      | A T .      |     |
| CTCACACCCG | ACCAATCAGA | GAGCTCACTA | AAATGCTAAT | CAGGCAAAAA | 700 |
| S H P      | T N Q R    | A H .      | N A N      | Q A K T    |     |
| H T R      | P I R      | E L T K    | M L I      | R Q K      |     |
| L T P D    | Q S E      | S S L      | K C .      | S G K N    |     |

11/27

FIG 5 (suite)

| 10         | 20         | 30         | 40         | 50         |      |
|------------|------------|------------|------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| CAGGAGGTAA | AGCAATAGCC | AATCATCIAT | TGCTGAGAG  | CACAGCGGA  | 750  |
| G G K      | A I A      | N H L L    | P E S      | T A G      |      |
| Q E V K    | Q . P      | I I Y      | C L R A    | Q R E      |      |
| R R .      | S N S Q    | S S I A    | . E H      | S G K      |      |
| AGGACAAGGA | TTGGGATATA | AATCAGGCA  | TTCAAGCCAG | CAACAGCAAC | 800  |
| R T R I    | G I .      | T Q A      | F K P A    | T A T      |      |
| G Q G      | L G Y K    | L R H      | S S Q      | Q Q Q P    |      |
| D K D      | W D I      | N S G I    | Q A S      | N S N      |      |
| CCCCTTGGG  | TCCCCGCCA  | TGTATGGGA  | GCTCTGTTT  | CACCTATTT  | 850  |
| P F G      | S P P I    | V W E      | L C F      | H S I S    |      |
| P L G      | P L P      | L Y G S    | S V F      | T L F      |      |
| P L W V    | P S H      | C M G      | A L F S    | L Y F      |      |
| CACCTATTA  | AATCATGCAA | CTGCACTCTT | CTGGTCCGIG | TTTTTATGG  | 900  |
| L Y .      | I M Q      | L H S S    | G P C      | F L W      |      |
| H S I K    | S C N      | C T L      | L V R V    | F Y G      |      |
| T L L      | N H A T    | A L F      | W S V      | F F M A    |      |
| CTCAAGCTGA | GCTTTTGTTC | GCCATCCACC | ACTGCTGTTT | GCCACCGICA | 950  |
| L K L S    | F C S      | P S T      | T A V C    | H R H      |      |
| S S .      | A F V R    | H P P      | L L F      | A T V T    |      |
| Q A E      | L L F      | A I H H    | C C L      | P P S      |      |
| CAGACCCGCT | GCTGACTTCC | ATCCCTTTGG | ATCCAGCAGA | GTTGCCACTG | 1000 |
| R P A      | A D F H    | P F G      | S S R      | V S T V    |      |
| D P L      | L T S      | I P L D    | P A E      | C P L      |      |
| Q T R C    | . L P      | S L W      | I Q Q S    | V H C      |      |
| TGCTCTGAT  | CCAGCGAGGT | ATCCATTGCC | ACTCCCGATC | AGGCTAAAGG | 1050 |
| L L I      | Q R G      | T H C H    | S R S      | G . R      |      |
| C S .      | S S E V    | P I A      | T P D Q    | A K G      |      |
| A P D      | P A R Y    | P L P      | L P I      | R L K A    |      |



12/27

FIG 5 (suite)

| 10         | 20         | 30         | 40          | 50         |      |
|------------|------------|------------|-------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890 |      |
| CTTGCCATTG | TTCCTGCATG | GCTAAGTGCC | TGGGTTTGTC  | CTAATAGAAC | 1100 |
| L A I V    | P A W      | L S A      | W V C P     | N R T      |      |
| L P L      | F L H G    | . V P      | G F V       | L I E L    |      |
| C H C      | S C M      | A K C L    | G L S       | . . N      |      |
|            |            |            |             |            |      |
| TGAACACTGG | TCACTGGGTT | CCATGGTTC  | CTTCCATGAC  | CCACGGCTTC | 1150 |
| E H W      | S L G S    | M V L      | F H D       | P R L L    |      |
| N T G      | H W V      | P W F S    | S M T       | H G F      |      |
| . T L V    | T G F      | H G S      | L P .       | P T A S    |      |
|            |            |            |             |            |      |
| TAATAGAGCT | ATAACACTCA | CCGCATGGCC | CAAGATTCCA  | TTCCTTGGIA | 1200 |
| I E L      | . H S      | P H G P    | R F H       | S L V      |      |
| . . S Y    | N T H      | R M A      | Q D S I     | P W Y      |      |
| N R A      | I T L T    | A W P      | K I P       | F L G I    |      |
|            |            |            |             |            |      |
| TCTGTGAGGC | CAAGAACCCC | AGGTCAGAGA | ANGTGAGGCT  | TGCCACCATT | 1250 |
| S V R P    | R T P      | G Q R      | X .         | G L P P F  |      |
| L . G      | Q E P Q    | V R E      | X E A       | C H H L    |      |
| C E A      | K N P      | R S E X    | V R L       | A T I      |      |
|            |            |            |             |            |      |
| TGGGAAGTGG | CCCACTGCCA | TTTTGGTAGC | GGCCCAACCAC | CATCTTGGGA | 1300 |
| G K W      | P T A I    | L V A      | A H H       | H L G S    |      |
| G S G      | P L P      | F W .      | R P T T     | I L G      |      |
| W E V A    | H C H      | F G S      | G P P P     | S W E      |      |
|            |            |            |             |            |      |
| GCTGTGGGAG | CAAGGATCCC | CCAGTAACA  |             |            | 1329 |
| C G S      | K D P      | P V T      |             |            |      |
| A V G A    | R I P      | Q .        |             |            |      |
| L W E      | Q G S P    | S N        |             |            |      |

13/27

FIG 6

| 10         | 20          | 30              | 40         | 50         |     |
|------------|-------------|-----------------|------------|------------|-----|
| 1234567890 | 1234567890  | 1234567890      | 1234567890 | 1234567890 |     |
| CCTAGAACGT | ATTCTGGAGA  | ATTGGGACCA      | ATGTGACACT | CAGACGCTAA | 50  |
| P R T Y    | S G E L     | G P M .         | H S D A K  |            |     |
| L E R I    | L E N W     | D Q C D T       | Q T L R    |            |     |
| . N V F    | W R I G T N | V T L R R .     |            |            |     |
| GAAAGAAAG  | ATTTATATTC  | TTCTGCAGTA      | COGCGTGGCC | ACAATATCTT | 100 |
| K E T I    | Y I L L     | Q Y R L A       | T I S S    |            |     |
| K K R F    | I F F C S T | A W P Q Y P     |            |            |     |
| E R N D    | L Y S S A V | P P G H N I L   |            |            |     |
| CTTCAAGGA  | GAGAAAGCTG  | GCTTCTGAG       | GGAAGTATAA | ATTATAACAT | 150 |
| S R E R    | N L A S .   | G K Y K L .     | H          |            |     |
| L Q G R    | E T W L P E | G S I N Y N I   |            |            |     |
| F K G E    | K P G F L R | E V . I I T S   |            |            |     |
| CATCTTACAG | CTAGACCTCT  | TCTGTAGAAA      | GGAGGGCAAA | TGGAGTGAAG | 200 |
| H L T A    | R P L L .   | K G G Q M E .   | S          |            |     |
| I L Q L    | D L F C R K | E G K W S E V   |            |            |     |
| S Y S .    | T S S V E R | R A N G V K     |            |            |     |
| TGCGATATGT | GCAAACCTTC  | TTTTCATTAA      | GAGACAACTC | ACAATTATGT | 250 |
| A I C A    | N F L F I K | R Q L T I M .   |            |            |     |
| P Y V Q    | T F F S L R | D N S Q L C     |            |            |     |
| C H M C    | K L S F H . | E T T H N Y V   |            |            |     |
| AAAAAGTGTG | GTTTATGCCC  | TACAGGAAGC      | CCTCAGAGTC | CACTTCCCTA | 300 |
| K V W F    | M P Y R K P | S E S T S L     |            |            |     |
| K K C G    | L C P T G S | P Q S P P P Y   |            |            |     |
| K S V V    | Y A L Q E A | L R V H L P T   |            |            |     |
| CCCCAGGTC  | COCTCCCGA   | CTCCTTCTC       | AACTAATAAG | GACCCCTT   | 350 |
| P Q R P    | L P D S F L | N . . G P P F   |            |            |     |
| P S V P    | S P T P S S | T N K D P P L   |            |            |     |
| P A S P    | P R L L P Q | L I R T P L     |            |            |     |
| TAAOCCAAAC | GGTCCAAAAG  | GAGATAGACA      | AAGGGGTAAA | CAATGAACCA | 400 |
| N P N G    | P K G D R Q | R G K Q . T K   |            |            |     |
| T Q T V    | Q K E I D K | G V N N E P     |            |            |     |
| . P K R    | S K R R .   | T K G . T M N Q |            |            |     |
| AAGAGTGOCA | ATATTCCCGG  | ATTATGCCCC      | CTCCAAGCAG | TGAGAGGAGG | 450 |
| E C Q Y    | S P I M P P | P S S E R R     |            |            |     |
| K S A N    | I P R L C P | L Q A V R G G   |            |            |     |
| R V P I    | F P D Y A P | S K Q . E E E   |            |            |     |
| AGAATTGGGC | CCAGCCAGAG  | TGCTGTACC       | TTTTCTCTC  | TCAGACTTAA | 500 |
| R I R P    | S Q S A C T | F F S L R L K   |            |            |     |
| E F G P    | A R V P V P | F S L S D L K   |            |            |     |
| N S A Q    | P E C L Y L | F L S Q T .     |            |            |     |

14/27

FIG 6 (suite)

| 10          | 20         | 30          | 40          | 50         |      |
|-------------|------------|-------------|-------------|------------|------|
| 1234567890  | 1234567890 | 1234567890  | 1234567890  | 1234567890 |      |
| AGCAAAATTAA | AATAGACCTA | GGTAAATTCT  | CAGATAACCC  | TGACGGCTAT | 550  |
| A N .       | N R P R    | . I L R .   | P . R L Y   |            |      |
| Q I K       | I D L      | G K F S     | D N P       | D G Y      |      |
| S K L K     | . T .      | V N S       | Q I T L     | T A I      |      |
| ATTGATGTTT  | TACAAGGGTT | AGGACAATCC  | TTTGATCTGA  | CATGGAGAGA | 600  |
| . C F       | T R V      | R T I L     | . S D       | M E R      |      |
| I D V L     | Q G L      | G Q S       | F D L T     | W R D      |      |
| L M F       | Y K G .    | D N P       | L I .       | H G E I    |      |
| TATAATGTTA  | CTACTAAATC | AGACACTAAC  | CCCAAAATGAG | AGAAGTGGCG | 650  |
| Y N V T     | T K S      | D T N       | P K .       | E K C R    |      |
| I M L       | L L N Q    | T L T       | P N E       | R S A A    |      |
| . C Y Y .   | I R H .    | P Q M R     | E V P       |            |      |
| CTGTAACTGC  | AGCCCGAGAG | TTTGGGGATC  | TTTGGTATCT  | CAGTCAGGCC | 700  |
| C N C       | S P R V    | W R S       | L V S       | Q S G Q    |      |
| V T A       | A R E      | F G D L     | W Y L       | S Q A      |      |
| L . L Q     | P E S      | L A I       | F G I S     | V R P      |      |
| AACAATAGGA  | TGACAACAGA | GGAAAGAACA  | ACTGCCACAG  | GCCAGCAGCC | 750  |
| Q . D       | D N R      | G K N N     | S H R       | P A G      |      |
| N N R M     | T T E      | E R T       | T P T G     | Q Q A      |      |
| T I G .     | Q Q R      | K E Q       | L P Q       | A S R Q    |      |
| AGTTCOCAGT  | GTAGACCTC  | ATTGGGACAC  | AGAATCAGAA  | CATGGAGATT | 800  |
| S S Q C     | R P S      | L G H       | R I R T     | W R L      |      |
| V P S       | V D P H    | W D T       | E S E       | H G D W    |      |
| F P V .     | T L        | I G T Q     | N Q N       | M E I      |      |
| GGTGCCACAA  | ACATTTCCTA | ACTTGGGTGC  | TAGAAGGACT  | GAGGAAAAT  | 850  |
| V P Q       | T F A N    | L R A       | R R T       | E E N .    |      |
| C H K       | H L L      | T C V L     | E G L       | R K T      |      |
| G A T N     | I C .      | L A C .     | K D .       | G K L      |      |
| AGGAAGAAGC  | CTATGAATTA | CTCAATGATG  | TOCACTATAA  | CACAGGGAAA | 900  |
| E E A       | Y E L      | L N D V     | H Y N       | T G K      |      |
| R K K P     | M N Y      | S M M       | S T I T     | Q G K      |      |
| G R S       | L .        | I T Q .     | C P L .     | H R E R    |      |
| GGAAGAAAAT  | CTTACTGCTT | TTCCTGGACAG | ACTAAGGGAG  | GCATTGACGA | 950  |
| G R K S     | Y C F      | S G Q       | T K G G     | I E E      |      |
| E E N       | L T A F    | L D R       | L R E       | A L R K    |      |
| K K I       | L L L      | F W T D     | . G R H .   | G          |      |
| AGCATACCTC  | CCTGTCACTT | GACTCTATTG  | AAGGCCAACT  | AATCTTAAAG | 1000 |
| A Y L       | P V T .    | L Y .       | R P T       | N L K G    |      |
| H T S       | L S P      | D S I E     | G Q L       | I L K      |      |
| S I P P     | C H L      | T L L       | K A N .     | S . R      |      |

15/27

FIG 6 (suite)

| 10            | 20              | 30         | 40         | 50         |      |
|---------------|-----------------|------------|------------|------------|------|
| 1234567890    | 1234567890      | 1234567890 | 1234567890 | 1234567890 |      |
| GATAAGTTTA    | TCACTCAGTC      | AGCTGCAGAC | ATTAGAAAAA | ACTTCAAAAG | 1050 |
| . V Y H S V   | S C R H         | . K K      | L Q K      |            |      |
| D K F I       | T Q S A A D     | I R K N    | F K S      |            |      |
| I S L S L S Q | L Q T L E K     | T S K V    |            |            |      |
| TCGCGCTTAG    | GCCCGGAGCA      | GAACCTAGAA | ACCCATTTTA | ACTTGGCATC | 1100 |
| S A L G       | P E Q N L E     | T L F N    | L A S      |            |      |
| L P .         | A R S R T .     | K P Y L    | T W H P    |            |      |
| C L R         | P G A E L R N   | P I .      | L G I      |            |      |
| CTCAGTTTTT    | TATAATAGAG      | ATCAGGAGGA | GCAGGCGAAA | CGGGACAAAC | 1150 |
| S V F Y N R D | Q E E Q A K     | R D K R    |            |            |      |
| Q F F I I E   | I R R S R R N   | G T N      |            |            |      |
| L S F L . .   | R S G G A G E T | G Q T      |            |            |      |
| GGGATAAAAA    | AAAAAGCGGG      | GGTCCACTAC | TTTAGTCATG | GCCCTCAGGC | 1200 |
| D K K K R G   | G P L L .       | S W P S G  |            |            |      |
| G I K K K G G | V H Y F S H G   | P Q A      |            |            |      |
| G . K K K G G | S T T L V M     | A L R Q    |            |            |      |
| AAGCAGACTT    | TGGAGGCTCT      | GCAAAAGGGA | AAAGCTGGGC | AAATCAAATG | 1250 |
| K Q T L E A L | Q K G K A G Q   | I K C      |            |            |      |
| S R L W R L C | K R E K L G     | K S N A    |            |            |      |
| A D F G G S   | A K G K S W A   | N Q M      |            |            |      |
| OCTAATAGGG    | CTGGCTTCCA      | GTGCGGCTTA | CAAGGACACT | TTAAAAAGA  | 1300 |
| L I G L A S S | A V Y K D T     | L K K I    |            |            |      |
| . . G W L P   | V R S T R T L   | . K R      |            |            |      |
| P N R A G F Q | C G L Q G H F   | K K D      |            |            |      |
| TTATCCAAGT    | AGAAATAAGC      | CGCCCCCTTG | TCCATGCCCC | TTAAGTCAAG | 1350 |
| I Q V E I S   | R P L V H A P   | Y V K      |            |            |      |
| L S K . K .   | A A P L S M P L | T S R      |            |            |      |
| Y P S R N K P | P P C P C P     | L R Q G    |            |            |      |
| GGATCACTG     | GAAGGCCAC       | TGCCCCAGGG | GATGAAGATA | CTCTGAGTCA | 1400 |
| G I T G R P T | A P G D E D T   | L S Q      |            |            |      |
| E S L E G P L | P Q G M K I     | L . V R    |            |            |      |
| N H W K A H   | C P R G .       | R Y S E S  |            |            |      |
| GAAGCCATTA    | ACCAGATGAT      | CCAGCAGCAG | GACTGAGGGT | GCCCCGGGGG | 1450 |
| K P L T R .   | S S S R T E G   | A R G E    |            |            |      |
| S H . P D D   | P A A G L R V   | P G A      |            |            |      |
| E A I N Q M I | Q Q Q D .       | G C P G R  |            |            |      |
| AGGGCCAGCC    | CATGCCATCA      | CCCTCACAGA | GCCCCGGGTA | TGTTTGACCA | 1500 |
| R Q P M P S   | P S' Q S P G Y  | V . P      |            |            |      |
| S A S P C H H | P H R A P G M   | F D H      |            |            |      |
| A P A H A I T | L T E P R V     | C L T I    |            |            |      |

16/27

## FIG 6 (suite)

| 10           | 20         | 30         | 40         | 50         |
|--------------|------------|------------|------------|------------|
| 1234567890   | 1234567890 | 1234567890 | 1234567890 | 1234567890 |
| TTGAGAGCCA A |            |            |            | 1511       |
| L            | R          | A          |            |            |
|              | E          | P          |            |            |
| E            | S          | Q          |            |            |

17/27

FIG 7

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATGGGCAGCA | GCCATCATCA | TCATCATCAC | AGCAGCGGCC | TGGTGCGCGG | 50  |
| M G S S    | H H H      | H H H      | S S G L    | V P R      |     |
| CGGCAGCCAT | ATGGCTAGCA | TGACTGGTGG | ACAGCAAATG | GGTCGGATCC | 100 |
| G S H      | M A S M    | T G G      | Q Q M      | G R I L    |     |
| TAGAAOGTAT | TCTGGAGAAT | TGGGACCAAT | GIGACACTCA | GACGCTAAGA | 150 |
| E R I      | L E N      | W D Q C    | D T Q      | T L R      |     |
| AAGAAACGAT | TTATATICTT | CTGCAGTAAC | GCCIGGCCAC | AATATCCTCT | 200 |
| K K R F    | I F F      | C S T      | A W P Q    | Y P L      |     |
| TCAAGGGAGA | GAAACCTGGC | TTCCTGAGGG | AAGTATAAAT | TATAACATCA | 250 |
| Q G R      | E T W L    | P E G      | S I N      | Y N I I    |     |
| TCTTACAGCT | AGACCTCTTC | TGTAGAAAGG | AGGGCAAATG | GAGTGAAGTG | 300 |
| L Q L      | D L F      | C R K E    | G K W      | S E V      |     |
| CCATATGTGC | AAACTTICTT | TTCATTAGA  | GACAACTCAC | AATTATGTAA | 350 |
| P Y V Q    | T F F      | S L R      | D N S Q    | L C K      |     |
| AAAGTGTGGT | TTATGCOCTA | CAGGAAGCCC | TCAGAGTCCA | CCTCCCTACC | 400 |
| K C G      | L C P T    | G S P      | Q S P      | P P Y P    |     |
| CCAGCGTCCC | CTCCCGACT  | CCTTCCTCAA | CTAATAAGGA | CCCCCTTTA  | 450 |
| S V P      | S P T      | P S S T    | N K D      | P P L      |     |
| ACCCAAACGG | TCCAAAAGGA | GATAGACAAA | GGGTAAACA  | ATGAACCAAA | 500 |
| T Q T V    | Q K E      | I D K      | G V N N    | E P K      |     |
| GAGTGCCAAT | ATTCCCCGAT | TATGCCCCCT | CCAAGCAGTG | AGAGGAGGAG | 550 |
| S A N      | I P R L    | C P L      | Q A V      | R G G E    |     |
| AATTGCGCCC | AGCCAGAGTG | CCTGTACCTT | TTTCTCTCTC | AGACTTAAAG | 600 |
| F G P      | A R V      | P V P F    | S L S      | D L K      |     |
| CAAATTAAAA | TAGACCTAGG | TAAATCTCTA | GATAACCTTG | ACGGCTATAT | 650 |
| Q I K I    | D L G      | K F S      | D N P D    | G Y I      |     |
| TGATGTTTTA | CAAGGGTTAG | GACAATCCTT | TGATCTGACA | TGGAGAGATA | 700 |
| D V L      | Q G L G    | Q S F      | D L T      | W R D I    |     |
| TAATGTTACT | ACTAAATCAG | ACACTAACCC | CAAATGAGAG | AAGTGCCGCT | 750 |
| M L L      | L N Q      | T L T P    | N E R      | S A A      |     |
| GTAAGTCAG  | CCCGAGAGTT | TGGGATCTTT | TGGTATCTCA | GTCAGGCCAA | 800 |
| V T A A    | R E F      | G D L      | W Y L S    | Q A N      |     |

18/27

## FIG 7 (suite)

| 10         | 20         | 30         | 40         | 50         |      |
|------------|------------|------------|------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| CAATAGGATG | ACAACAGAGG | AAAGAACAAC | TCCCACAGGC | CAGCAGGCAG | 850  |
| N R M      | T T E E    | R T T      | P T G      | Q Q A V    |      |
| TTCCCAGTGT | AGACCTCAT  | TGGACACAG  | AATCAGAACA | TGGAGATTGG | 900  |
| P S V      | D P H      | W D T E    | S E H      | G D W      |      |
| TGCCACAAAC | ATTIGCTAAC | TTGGTGCTA  | GAAGGACTGA | GGAAAACTAG | 950  |
| C H K H    | L L T      | C V L      | E G L R    | K T R      |      |
| GAAGAAGCCT | ATGAATTACT | CAATGATGTC | CACTATAACA | CAGGGAAAGG | 1000 |
| K K P      | M N Y S    | M M S      | T I T      | Q G K E    |      |
| AAGAAAATCT | TACTGCTTTT | CTGGACAGAC | TAAGGGAGGC | ATTGAGGAAG | 1050 |
| E N L      | T A F      | L D R L    | R E A      | L R K      |      |
| CATACTCCC  | TGTCACCTGA | CTCTATTGAA | GGCCAACTAA | TCTTAAAGGA | 1100 |
| H T S L    | S P D      | S I E      | G Q L I    | L K D      |      |
| TAAGTTTATC | ACTCAGTCAG | CTGCAGACAT | TAGAAAAAAC | TTCAAAAGTC | 1150 |
| K F I      | T Q S A    | A D I      | R K N      | F K S L    |      |
| TGCTTAAGCT | TGCGGCGGCA | CTCGAGCACC | ACCACCACCA | CCACTGAGAT | 1200 |
| P K L      | A A A      | L E H H    | H H H      | H . D      |      |
| COGGCTGCTA | ACAAAGCCCC | AAAGGAAGCT | GAGTTGGCTN | GTGGCNA    | 1247 |
| P A A N    | K A R      | K E A      | E L A X    | G          |      |

19/27

FIG 8

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATGGCTAGCA | TGACTGGTGG | ACAGCAAATG | GGTCGGATCC | TAGAACTAT  | 50  |
| M A S M    | T G G      | Q Q M      | G R I L    | E R I      |     |
| TCTGGAGAAT | TGGGACCAAT | GTGACACTCA | GACGCTAAGA | AAGAAACGAT | 100 |
| L E N      | W D Q C    | D T Q      | T L R      | K K R F    |     |
| TTATATTCTT | CTGCAGTACC | GCCTGGCCAC | AATATCCTCT | TCAAGGGAGA | 150 |
| I F F      | C S T      | A W P Q    | Y P L      | Q G R      |     |
| GAAACCTGGC | TTCCTGAGGG | AAGTATAAAT | TATAACATCA | TCTTACAGCT | 200 |
| E T W L    | P E G      | S I N      | Y N I I    | L Q L      |     |
| AGACCTCTTC | TGTAGAAAGG | AGGGCAAATG | GAGTGAAGTG | CCATATGTGC | 250 |
| D L F      | C R K E    | G K W      | S E V      | P Y V Q    |     |
| AAACTTTCTT | TTCATTAGA  | GACAACTCAC | AATTATGTAA | AAAGTGTGGT | 300 |
| T F F      | S L R      | D N S Q    | L C K      | K C G      |     |
| TTATGCCCTA | CAGGAAGCCC | TCAGAGTCCA | CCTCCCTACC | CCAGGGTCCC | 350 |
| L C P T    | G S P      | Q S P      | P P Y P    | S V P      |     |
| CTCCCCGACT | CCTTCTCTCA | CTAATAAGGA | CCCCCCTTTA | ACCCAAAAGG | 400 |
| S P T      | P S S T    | N K D      | P P L      | T Q T V    |     |
| TOCAAAAGGA | GATAGACAAA | GGGGTAAACA | ATGAACCAAA | GAGTGCCAAT | 450 |
| Q K E      | I D K      | G V N N    | E P K      | S A N      |     |
| ATTCCCCGAT | TATGCCCCCT | CCAAGCAGTG | AGAGGAGGAG | AATTGGGCCC | 500 |
| I P R L    | C P L      | Q A V      | R G G E    | F G P      |     |
| AGCCAGAGTG | CCTGTACCTT | TTTCTCTCTC | AGACTTAAAG | CAAATTAAAA | 550 |
| A R V      | P V P F    | S L S      | D L K      | Q I K I    |     |
| TAGACCTAGG | TAAATTCTCA | GATAACCTTG | ACGGCTATAT | TGATGTTTAA | 600 |
| D L G      | K F S      | D N P D    | G Y I      | D V L      |     |
| CAAGGGTTAG | GACAATCCTT | TGATCTGACA | TGGAGAGATA | TAATGTTACT | 650 |
| Q G L G    | Q S F      | D L T      | W R D I    | M L L      |     |
| ACTAAATCAG | AACTAAACCC | CAAATGAGAG | AAGTGGCGCT | GTAAGTGCAG | 700 |
| L N Q      | T L T P    | N E R      | S A A      | V T A A    |     |
| CCCGAGAGTT | TGGGATCTTT | TGGTATCTCA | GTCAGGCCAA | CAATAGGATG | 750 |
| R E F      | G D L      | W Y L S    | Q A N      | N R M      |     |
| ACAACAGAGG | AAAGAACAAC | TOCCACAGGC | CAGCAGGCAG | TTOCCAGTGT | 800 |
| T T E E    | R T T      | P T G      | Q Q A V    | P S V      |     |



## FIG 8 (suite)

| 10          | 20         | 30         | 40         | 50         |      |
|-------------|------------|------------|------------|------------|------|
| 1234567890  | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| AGACCCATCAT | TGGGACACAG | AATCAGAACA | TGGAGATTGG | TGCCACAAAC | 850  |
| D P H       | W D T E    | S E H      | G D W      | C H K H    |      |
| ATTTGCTAAC  | TTGGTGCTA  | GAAGGACTGA | CGAAACTAG  | GAAGAAGCCT | 900  |
| L L T       | C V L      | E G L R    | K T R      | K K P      |      |
| ATGAATTACT  | CAATGATGTC | CACTATAACA | CAGGGAAGG  | AAGAAAATCT | 950  |
| M N Y S     | M M S      | T I T      | Q G K E    | E N L      |      |
| TACTGCTTTT  | CTGGACAGAC | TAAGGGAGGC | ATTGAGGAAG | CATACCTCCC | 1000 |
| T A F       | L D R L    | R E A      | L R K      | H T S L    |      |
| TGTCACCTGA  | CTCTATTGAA | GGCCAACATA | TCTTAAAGGA | TAAGTTTATC | 1050 |
| S P D       | S I E      | G Q L I    | L K D      | K F I      |      |
| ACTCAGTCAG  | CTGCAGACAT | TAGAAAAAAC | TTCAAAAGTC | TGCCTAAGCT | 1100 |
| T Q S A     | A D I      | R K N      | F K S L    | P K L      |      |
| TGCGGCGCA   | CTGAGCACC  | ACCACCAACA | CCACTGAGAT | CCGGCTGCTA | 1150 |
| A A A       | L E H H    | H H H      | H .        | D P A A N  |      |
| ACAAAGCCCG  | AAAGGAAGCT | GAGTTGGCTG | GTGCA      |            | 1186 |
| K A R       | K E A      | E L A      | G G        |            |      |

FIG 9

| 10         | 20         | 30          | 40          | 50          |     |
|------------|------------|-------------|-------------|-------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890  | 1234567890  |     |
| TGTCGCTGT  | GCTCCTGATC | CAGCACAGGC  | GCCCATTTGCC | TCTCCCAATT  | 50  |
| C P L C    | S . S      | S T G       | A H C L     | S Q L       |     |
| V R C      | A P D P    | A Q A       | P I A       | S P N W     |     |
| S A V      | L L I      | Q H R R     | P L P       | L P I       |     |
| GGCTAAAGG  | CTTGCCATTG | TTCCTGCACA  | GCTAAGTGCC  | TGGGTTCATC  | 100 |
| G . R      | L A I V    | P A Q       | L S A       | W V H P     |     |
| A K G      | L P L      | F L H S     | . V P       | G F I       |     |
| G L K A    | C H C      | S C T       | A K C L     | G S S       |     |
| CTAATCGAGC | TGAACACTAG | TCACTGGGTT  | CCACGGTTCT  | CTTCCATGAC  | 150 |
| N R A      | E H .      | S L G S     | T V L       | F H D       |     |
| L I E L    | N T S      | H W V       | P R F S     | S M T       |     |
| . S S      | . T L V    | T G F       | H G S       | L P . P     |     |
| CCATGGCTTC | TAATAGAGCT | ATAACACTCA  | CTGCATGGTC  | CAAGATTCCA  | 200 |
| P W L L    | I E L      | . H S       | L H G P     | R F H       |     |
| H G F      | . . S Y    | N T H       | C M V       | Q D S I     |     |
| M A S      | N R A      | I T L T     | A W S       | K I P       |     |
| TTCCTTGGA  | TCCGTGAGAC | CAAGAACCCC  | AGGTCAGAGA  | ACACAAGGCT  | 250 |
| S L E      | S V R P    | R T P       | G Q R       | T Q G L     |     |
| P W N      | P . D      | Q E P Q     | V R E       | H K A       |     |
| F L G I    | R E T      | K N P       | R S E N     | T R L       |     |
| TGCCACCATG | TTGGAAGCAG | CCCAACCACCA | TTTTGGAAGC  | AGCCCCGCCAC | 300 |
| P P C      | W K Q      | P T T I     | L E A       | A R H       |     |
| C H H V    | G S S      | P P P       | F W K Q     | P A T       |     |
| A T M      | L E A A    | H H H       | F G S       | S P P L     |     |
| TATCTTGGGA | GCTCTGGGAG | CAAGGACCCC  | AGGTAAACAAT | TTGGTGACCA  | 350 |
| Y L G S    | S G S      | K D P       | R . Q F     | G D H       |     |
| I L G      | A L G A    | R T P       | G N N       | L V T T     |     |
| S W E      | L W E      | Q G P Q     | V T I       | W . P       |     |
| CGAAGGGACC | TGAATCCGCA | ACCATGAAGG  | GATCTCCAAA  | GCAATTGGAA  | 400 |
| E G T      | . I R N    | H E G       | I S K       | A I G N     |     |
| K G P      | E S A      | T M K G     | S P K       | Q L E       |     |
| R R D L    | N P Q      | P . R       | D L Q S     | N W K       |     |

22/27

FIG 9 (suite)

| 10         | 20         | 30          | 40         | 50         |     |
|------------|------------|-------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890 | 1234567890 |     |
| ATGTTCTCTC | CAAGGCAAAA | ATGCCCCCTAA | GATGTATTCT | GGAGAATTGG | 450 |
| V P P      | K A K      | M P L R     | C I L      | E N W      |     |
| M F L P    | R Q K      | C P .       | D V F W    | R I G      |     |
| C S S      | Q G K N    | A P K       | M Y S      | G E L G    |     |
| GACCAATTTG | ACCTTCAGAC | AGTAAGAAAA  | AAATGACTTA | TATTCTTCTG | 500 |
| D Q F D    | P Q T      | V R K       | K .        | L I F F C  |     |
| T N L      | T L R Q    | . E K       | N D L      | Y S S A    |     |
| P I .      | P S D      | S K K K     | M T Y      | I L L      |     |
| CAGTACCGCC | CTGGOCACGA | TATCCTCTTC  | AAGGGGAGA  | AACTTGGCCT | 550 |
| S T A      | L A T I    | S S S       | R G R      | N L A S    |     |
| V P P      | W P R      | Y P L Q     | G G E      | T W P      |     |
| Q Y R P    | G H D      | I L F       | K G E K    | P G L      |     |
| CCTGAGGGAA | GTATAAATTA | TAACACCATC  | TTACAGCTAG | ACCTGTTTIG | 600 |
| . G K      | Y K L      | . H H L     | T A R      | P V L      |     |
| P E G S    | I N Y      | N T I       | L Q L D    | L F C      |     |
| L R E V    | . I I      | T P S       | Y S .      | T C F V    |     |
| TAGAAAAGGA | GGCAAATGGA | GTGAAGTGCC  | ATATTTACAA | ACTTCTTTT  | 650 |
| . K R R    | Q M E      | . S A       | I F T N    | F L F      |     |
| R K G      | G K W S    | E V P       | Y L Q      | T F F S    |     |
| E K E      | A N G      | V K C H     | I Y K      | L S F      |     |
| CATTAAAAGA | CAACTCGCAA | TTATGTTAAC  | AGTGIGATT  | TGTTCTCTAC | 700 |
| I K R      | Q L A I    | M L T       | V .        | F V F L H  |     |
| L K D      | N S Q      | L C .       | Q C D L    | C S Y      |     |
| H . K T    | T R N      | Y V N       | S V I C    | V P T      |     |
| ACGGAAGCCC | TCAGATTCTA | CTCCCCACCC  | CCGGCATCTC | CCCTGAATCC | 750 |
| G S P      | Q I L      | L P T P     | G I S      | P E S      |     |
| T E A L    | R F Y      | S P P       | P A S P    | L N P      |     |
| R K P      | S D S T    | P H P       | R H L      | P . I P    |     |
| CTCCCCAACT | TATT       |             |            |            | 764 |
| L P N L    |            |             |            |            |     |
| S P T Y    |            |             |            |            |     |
| P Q L I    |            |             |            |            |     |

FIG 10

| 10              | 20            | 30          | 40          | 50         |     |
|-----------------|---------------|-------------|-------------|------------|-----|
| 1234567890      | 1234567890    | 1234567890  | 1234567890  | 1234567890 |     |
| TGTCGGCTGT      | GCTCTGATC     | CAGCACAGGC  | GCCCATGTGCC | TCTCCCAATT | 50  |
| C P L C S . S   | S T G A H C L | S Q L       |             |            |     |
| V R C A P D P   | A Q A P I A   | S P N W     |             |            |     |
| S A V L L I     | Q H R R       | P L P L P I |             |            |     |
| GGGCTAAAGG      | CTTGCCATTG    | TTCTTGCACA  | GCTAAGTGCC  | TGGGTTCATC | 100 |
| G . R L A I V   | P A Q L S A   | W V H P     |             |            |     |
| A K G L P L     | F L H S . V P | G F I       |             |            |     |
| G L K A C H C   | S C T A K C L | G S S       |             |            |     |
| CTAATCGAGC      | TGAACACTAG    | TCACTGGGTT  | CCACGGTTCT  | CTTCCATGAC | 150 |
| N R A E H .     | S L G S T V L | F H D       |             |            |     |
| L I E L N T S   | H W V P R F S | S M T       |             |            |     |
| . S S . T L V   | T G F H G S   | L P . P     |             |            |     |
| CCATGGCTTC      | TAATAGAGCT    | ATAACACTCA  | CTGCATGGTC  | CAAGATTCCA | 200 |
| P W L L I E L . | H S L H G P   | R F H       |             |            |     |
| H G F . . S Y   | N T H C M V   | Q D S I     |             |            |     |
| M A S N R A     | I T L T A W S | K I P       |             |            |     |
| TTCTTTGGAA      | TCGGTGAGAC    | CAAGTACCCC  | AGGTCAGAGA  | ACACAAGGCT | 250 |
| S L E S V R P   | R T P G Q R   | T Q G L     |             |            |     |
| P W N P . D     | Q E P Q V R E | H K A       |             |            |     |
| F L G I R E T   | K N P R S E N | T R L       |             |            |     |
| TGCCACCATG      | TTGGAAGCAG    | CCCACCACCA  | TTTGTGAAGC  | GGCCCGCCAC | 300 |
| P P C W K Q     | P T T I L E A | A R H       |             |            |     |
| C H H V G S S   | P P P F W K R | P A T       |             |            |     |
| A T M L E A A   | H H H F G S   | G P P L     |             |            |     |
| TATCTTGGGA      | GCTCTGGGAG    | CAAGGACCCC  | CAGGTAACAA  | TTTGGTGACC | 350 |
| Y L G S S G S   | K D P Q V T I | W . P       |             |            |     |
| I L G A L G A   | R T P R . Q   | F G D H     |             |            |     |
| S W E L W E     | Q G P P G N N | L V T       |             |            |     |
| ACGAAGGGAC      | CTGAATCCGC    | AACCATGAAG  | GGATCTCCAA  | AGCAATTGGA | 400 |
| R R D L N P Q   | P . R D L Q   | S N W K     |             |            |     |
| E G T . I R     | N H E G I S K | A I G       |             |            |     |
| T K G P E S A   | T M K G S P K | Q L E       |             |            |     |

FIG 10 (suite)

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| AATGTTCTC  | CCAAGGCAA  | AATGCCCCA  | AGATGTATC  | TGGAGAATTG | 450 |
| C S S      | Q G K      | N A P K    | M Y S      | G E L      |     |
| N V P P    | K A K      | M P L      | R C I L    | E N W      |     |
| M F L      | P R Q K    | C P .      | D V F      | W R I G    |     |
| GGACCAATCT | GACCTCAGA  | CAGTAAGAA  | AAAAATGACT | TATATCTTC  | 500 |
| G P I .    | P S D      | S K K      | K N D L    | Y S S      |     |
| D Q S      | D P Q T    | V R K      | K M T      | Y I L L    |     |
| T N L      | T L R Q    | . E K      | K . L      | I F F      |     |
| TGCAGTACCG | CCTGGCCACG | GATATCTCT  | TCAAGGGGA  | GAAACCTGGC | 550 |
| A V P      | P G H G    | Y P L      | Q G G      | E T W P    |     |
| Q Y R      | L A T      | D I L F    | K G E      | K P G      |     |
| C S T A    | W P R      | I S S      | S R G R    | N L A      |     |
| CTCCTGAGGG | AAGTATAAAT | TATAACACCA | TCTTACAGCT | AGACCTGTTT | 600 |
| P E G      | S I N      | Y N T I    | L Q L      | D L F      |     |
| L L R E    | V . I      | I T P      | S Y S .    | T C F      |     |
| S . G      | K Y K L    | . H H      | L T A      | R P V L    |     |
| TGTAGAAAAG | GAGGCAAATG | GAGTGAAGTG | CCATATTAC  | AACTTTCTT  | 650 |
| C R K G    | G K W      | S E V      | P Y L Q    | T F F      |     |
| V E K      | E A N G    | V K C      | H I Y      | K L S F    |     |
| . K R      | R Q M      | E . S A    | I F T      | N F L      |     |
| TTCATTAAAA | GACAACTCGC | AATTATGTAA | ACAGTGTGAT | TTGTGTCCTA | 700 |
| S L K      | D N S Q    | L C K      | Q C D      | L C P T    |     |
| H . K      | T T R      | N Y V N    | S V I      | C V L      |     |
| F I K R    | Q L A      | I M .      | T V .      | F V S Y    |     |
| CAGGAAGCCC | TCAGATCTAC | CTCCCTACCC | CGGCATCTCC | CTGACTCCTT | 750 |
| G S P      | Q I Y      | L P T P    | A S P .    | L L        |     |
| Q E A L    | R S T      | S L P      | R H L P    | D S F      |     |
| R K P      | S D L P    | P Y P      | G I S      | L T P S    |     |
| CCCCAACTAA | TAAGGACCCA | CTTCAGCCCA | AACAGTCCAA | AAGGACATAG | 800 |
| P Q L I    | R T H      | F S P      | N S P K    | G H        |     |
| P N .      | . G P T    | S A Q      | T V Q      | K D I      |     |
| P T N      | K D P      | L Q P K    | Q S K      | R T .      |     |

FIG 11

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| GGCATTGATA | GCACCCATCA | GATGGCCAAA | TCATTATTIA | CTGGACCAGG | 50  |
| G I D S    | T H Q      | M A K      | S L F T    | G P G      |     |
| A L I      | A P I R    | W P N      | H Y L      | L D Q A    |     |
| H . .      | H P S      | D G Q I    | I I Y      | W T R      |     |
| CCTTTTCAAA | ACTATCAAGC | AGATAGGGCC | CGTGAAGCAT | GCCAAAGAAA | 100 |
| L F K      | T I K Q    | I G P      | V K H      | A K E I    |     |
| F S K      | L S S R    | . G P      | . S M      | P K K      |     |
| P F Q N    | Y Q A      | D R A      | R E A C    | Q R N      |     |
| TAATCCCCIG | CCTTATGGCC | ATGTTCTTTC | AGGAGAACAA | AGAACAGGCC | 150 |
| I P C      | L I A      | M F L Q    | E N K      | E Q A      |     |
| . S P A    | L S P      | C S F      | R R T K    | N R P      |     |
| N P L      | P Y R H    | V P S      | G E Q      | R T G H    |     |
| ATTACCCAGG | GGAAGACTGG | CAACTAGATT | TTACCCACAT | GGCCAAATGT | 200 |
| I T Q G    | K T G N    | . I        | L P T W    | P N V      |     |
| L P R      | G R L A    | T R F      | Y P H      | G Q M S    |     |
| Y P G      | E D W      | Q L D F    | T H M      | A K C      |     |
| CAGGGATTTC | AGCATCTACT | AGTCTGGGCA | GATACTTTCA | CTGGITGGGT | 250 |
| R D F      | S I Y .    | S G Q      | I L S      | L V G W    |     |
| G I S      | A S T      | S L G R    | Y F H      | W L G      |     |
| Q G F Q    | H L L      | V W A      | D T F T    | G W V      |     |
| GGAGICTTCT | OCTTGTAGGA | CAGAAAAGAC | CCAAGAGGTA | ATAAAGGCAC | 300 |
| S L L      | L V G      | Q K R P    | K R .      | . R H      |     |
| G V F S    | L . D      | R K D      | P R G N    | K G T      |     |
| E S S      | P C R T    | E K T      | Q E V      | I K A L    |     |
| TAATGAATA  | ATTCCAGAT  | TTGGACTTCC | CCAGGATTIA | CAGGGTGACA | 350 |
| . . N N    | S Q I      | W T S      | P R I T    | G . Q      |     |
| N E I      | I P R F    | G L P      | P G L      | Q G D N    |     |
| M K .      | F P D      | L D F P    | Q D Y      | R V T      |     |

FIG 11 (suite)

| 10         | 20         | 30          | 40         | 50         |     |
|------------|------------|-------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890 | 1234567890 |     |
| ATGGCCCCGC | TTTCAAGGCT | GCAGTAACCC  | AGGGAGTATC | CCAGGTGTTA | 400 |
| W P R      | F Q G C    | S N P       | G S I      | P G V R    |     |
| G P A      | F K A      | A V T Q     | G V S      | Q V L      |     |
| M A P L    | S R L      | Q . P       | R E Y P    | R C .      |     |
| GGCATACAAT | ATCACTTACA | CTGTGCGCTGG | AGGCCACAAT | CCTCCAGAAA | 450 |
| H T I      | S L T      | L C L E     | A T I      | L Q K      |     |
| G I Q Y    | H L H      | C A W       | R P Q S    | S R K      |     |
| A Y N      | I T Y T    | V P G       | G H N      | P P E K    |     |
| AGTCAAGAAA | ATGAATGAAA | CACTCAAAGA  | TCTAAAAAAG | CTAACCCAAG | 500 |
| S Q E N    | E . N      | T Q R       | S K K A    | N P R      |     |
| V K K      | M N E T    | L K D       | L K K      | L T Q E    |     |
| S R K      | . M K      | H S K I     | . K S      | . P K      |     |
| AAACCCACAT | TGCATGACCT | GTTCGTGTGC  | CTATAACCTT | ACTAAGAATC | 550 |
| N P H      | C M T C    | S V A       | Y N L      | T K N P    |     |
| T H I      | A . P      | V L L P     | I T L      | L R I      |     |
| K P T L    | H D L      | F C C       | L . P Y    | . E S      |     |
| CATAACTATC | CCCCAAAAG  | CAGGACTTAG  | CCCATACGAG | ATGCTATATG | 600 |
| . L S      | P K K      | Q D L A     | H T R      | C Y M      |     |
| H N Y P    | P K S      | R T .       | P I R D    | A I W      |     |
| I T I      | P Q K A    | G L S       | P Y E      | M L Y G    |     |
| GATGGCCTTT | CCTAACCAAT | GACCTTGTGC  | TTGACTGAGA | AATGGCCAAC | 650 |
| D G L S    | . P M      | T L C       | L T E K    | W P T      |     |
| M A F      | P N Q .    | P C A       | . L R      | N G Q L    |     |
| W P F      | L T N      | D L V L     | D . E      | M A N      |     |
| TTAGTTGCAG | ACATCACCTC | CTTAGCCAAA  | TATCAACAAG | TTCTTAAAAC | 700 |
| . L Q      | T S P P    | . P N       | I N K      | F L K H    |     |
| S C R      | H H L      | L S Q I     | S T S      | S . N      |     |
| L V A D    | I T S      | L A K       | Y Q Q V    | L K T      |     |

## FIG 11 (suite)

---

| 10         | 20         | 30         | 40          | 50          |     |
|------------|------------|------------|-------------|-------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890  |     |
| ATCACAGGGA | ACCTGTCCCC | GAGAGGAGGG | AAAGGAAC TA | TTCCACCC TG | 750 |
| H R E      | P V P      | E R R E    | R N Y       | S T L       |     |
| I T G N    | L S P      | R G G      | K G T I     | P P W       |     |
| S Q G      | T C P R    | E E G      | K E L       | F H P G     |     |

GTGACATG

758

V T

. H

D M



INSTITUT NATIONAL

de la

PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE  
PRELIMINAIREétabli sur la base des dernières revendications  
déposées avant le commencement de la rechercheN° d'enregistrement  
nationalFA 544515  
FR 9708816

| DOCUMENTS CONSIDERES COMME PERTINENTS   |  | Revendications<br>concernées<br>de la demande<br>examinée |
|---|--|---|
| Catégorie   | Citation du document avec indication, en cas de besoin,<br>des parties pertinentes   |   |
| X   | DATABASE EMBLEMEST11<br>SEQ ID HSZZ32436 AC AA327384<br>EST30710 Colon I Homo sapiens cDNA 5',<br>18 avril 1997<br>M.D. ADAMS ET AL.: "Initial assessment of<br>human gene diversity and expression<br>patterns based upon 83 million nucleotides<br>of cDNA sequence"<br>XP002059257<br>& NATURE.,<br>vol. 377supp, 28 septembre 1995, pages<br>3-174,<br>LONDON GB | 1,2,<br>14-20   |
| D,X   | FR 2 737 500 A (BIO MERIEUX S.A.)<br>7 février 1997<br><br>* le document en entier *   | 1-3,5,<br>13-20,<br>22-26                                 |
| D,X   | WO 95 21256 A (BIO MERIEUX S.A.)<br>10 août 1995<br><br>* le document en entier *  | 1-3,5,<br>13-20,<br>22-26                                 |
| X   | G. LA MANTIA ET AL.: "Identification and<br>characterization of novel human endogenous<br>retroviral sequences preferentially<br>expressed in undifferentiated embryonal<br>carcinoma cells"<br>NUCLEIC ACIDS RESEARCH,<br>vol. 19, no. 7, 1991, pages 1513-1520,<br>XP002059255<br>OXFORD GB<br>* le document en entier *   | 1-3,5,<br>13-17,<br>22-24                                 |
| -/--  |  |   |
| Date d'achèvement de la recherche   |  | Examineur   |
| 6 octobre 1998  |  | Cupido, M   |
| CATEGORIE DES DOCUMENTS CITES   |  |   |
| X : particulièrement pertinent à lui seul<br>Y : particulièrement pertinent en combinaison avec un<br>autre document de la même catégorie<br>A : pertinent à l'encontre d'au moins une revendication<br>ou arrière-plan technologique général<br>O : divulgation non-écrite<br>P : document intercalaire                                  |  |   |
| T : théorie ou principe à la base de l'invention<br>E : document de brevet bénéficiant d'une date antérieure<br>à la date de dépôt et qui n'a été publié qu'à cette date<br>de dépôt ou qu'à une date postérieure.<br>D : cité dans la demande<br>L : cité pour d'autres raisons<br>& : membre de la même famille, document correspondant |  |   |

INSTITUT NATIONAL

de la

PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE  
PRELIMINAIREétabli sur la base des dernières revendications  
déposées avant le commencement de la rechercheN° d'enregistrement  
nationalFA 544515  
FR 9708816

| DOCUMENTS CONSIDERES COMME PERTINENTS   |  | Revendications concernées de la demande examinée |
|---|--|--|
| Catégorie   | Citation du document avec indication, en cas de besoin, des parties pertinentes  |  |
| T   | H. PERRON ET AL.: "Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis"<br>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,<br>vol. 94, no. 14, 8 juillet 1997, pages 7538-7588, XP002059256<br>WASHINGTON US<br>* le document en entier * | 1-3,5,<br>13-20,<br>22-26                        |
| A   | WO 94 28138 A (UNIVERSITY COLLEGE LONDON)<br>8 décembre 1994<br>* le document en entier *  | 1-3,5,<br>13-20,<br>22-26                        |
| A   | WO 93 07259 A (SCLEROSE-FORENINGEN)<br>15 avril 1993<br>* le document en entier *  | 1-3,5,<br>13-20,<br>22-26                        |
|   |  | DOMAINES TECHNIQUES<br>RECHERCHES (Int.CL.6)     |
|   |  |  |
| Date d'achèvement de la recherche   |  | Examineur  |
| 6 octobre 1998  |  | Cupido, M  |
| CATEGORIE DES DOCUMENTS CITES   |  |  |
| X : particulièrement pertinent à lui seul<br>Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie<br>A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général<br>O : divulgation non-écrite<br>P : document intercalaire                                  |  |  |
| T : théorie ou principe à la base de l'invention<br>E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.<br>D : cité dans la demande<br>L : cité pour d'autres raisons<br>& : membre de la même famille, document correspondant |  |  |

2

EPO FORM 1503 03.82 (P04C13)